"Sapienza" University of Rome

Protein Sequencing Strategy in NanoTechnology by Classical and Quantum Atomistic Models

XXXI

Ph.D. student: *Aldo Eugenio Rossini*

Mathematical Models for Engineering, Electromagnetics and Nanosciences

Supervisor: *Prof. Giuseppe Zollo* Tutor: *Dott. Fabrizio Gala*

Contents

List of Figures

Introduction

This thesis focuses on simulations for peptides sequencing, based on the Nanopore technology. Nanopore sensing is a single-molecule technique capable of detecting peptide and protein by monitoring the change in current generated by the interaction of protein with nanopore devices. To better understand what happens in the nanopore devices, we analyze them by means of two methods: one classical to study the ionic current and take an overview of the phenomena with the possibility of simulating more atoms; and one quantum to study the tunneling current and have a more precise signal.

For the classical model, the idea of using nanopores inserted in lipid membranes, as a tool for the analysis of single molecules, was inspired by the very intense molecular transport activity between the intracellular and the extracellular media as well as between different cellular organelles. Molecular transport across the naturally impermeable membranous structures of cells occurs through a variety of protein channels incorporated into the fluid mosaic of the lipid bilayers. These channels or pores act as gates through which a wide variety of molecules such as ions, sugars, nucleic acids and proteins can pass during their transport from one organelle to another, or from the cytoplasm to the outside of the cell. The ability of the pores to allow the passage of ions and larger molecules suggested that the ions could be used to carry an electric current, which in turn could drive larger polar molecules through the channels. The change in the ionic flow through the channel due to the transit of the macromolecule could depend on the structure of the particular translocating molecule.

The quantum study focuses on the ab-initio simulations of charge transport through nano-gap of graphene. The Latin word ab-initio ("from beginning") means that no adjustable parameters are needed in the simulation. An atomistic description of materials and nanosystems without fitting parameters has an unprecedented predictive power for the study of their physical properties and gives access to an extremely broad range of phenomena otherwise not observable. The ab-initio simulations applied for charge transport mechanisms have recently emerged in both pure and applied research, as a

powerful technique to gain insight into the quantum phenomena governing the conduction properties of systems with reduced dimensionality where classical charge transport models are no longer valid. One of the most widely used techniques for the ab-initio calculation of the electronic transport properties is based on the combination of density functional theory (DFT) for the electronic structure with the non-equilibrium Green's function (NEGF) for quantum transport.

But why do we bother studying into these new sequencing techniques? The first attempt to sequencing the human genome was started in 1990 by the US government under the name of the 'Human Genome Project'. This project was successfully completed in the year 2003, with a cost of about $ 2.7 billion; at that time, the so-called Sanger method [1] was used, now the DNA sequencing is carried out with nanopore devices 1 [2]. After DNA sequencing was born "Human proteome project" for the sequencing of proteins. The recents progresses in biology have allowed to select amino-acid sequences; but sequencing of amino-acid chains (peptides, proteins) in an effective manner (rapid, selective) is not possible yet. The importance of identifying the sequence and the structure of proteins and polypeptides is crucial, because mis-folding of proteins (structured and not), is believed to be related to neurodegenerative diseases such as Parkinson's and Alzheimer [3] as well as the post-translational modifications [4]. In order to achieve a better understanding of genetics, mutations, it is vital to have a low-cost, highly parallelizable way of sequencing. This could lead to so-called personalized medicine, meaning that it would be affordable for a person to have their genome sequenced in order to find potential weaknesses or predispositions for certain illnesses.

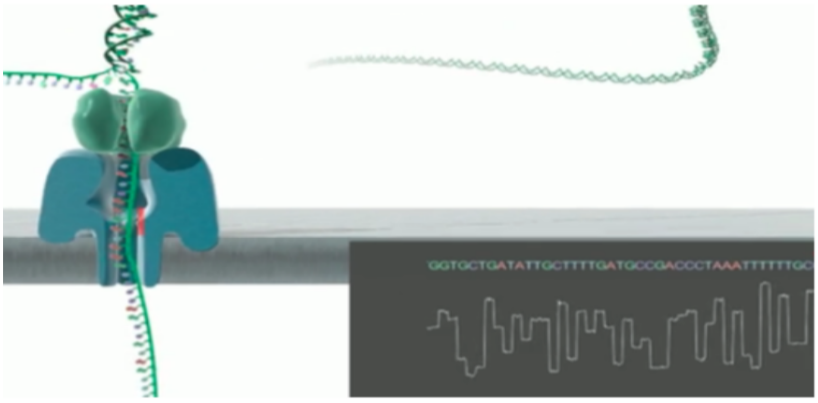


Figure 1: *Illustration of a single-stranded DNA molecule passing through αHL nanopore [2].*

Chapter 1

State of the Art

Initially intended as a tool for DNA and RNA analysis with the long term goal of rapid nucleic acid sequencing, the method of sensing with nanopores was later extended for investigating a broad spectrum of molecules ranging from metal ions, small organic compunds and short peptides to chemical warfare agents, proteins and bio-molecular complexes. In this chapter, the most significant advances in nucleic acid as well as in peptide and protein detection with nanopores will be presented.

1.1 Nanopore Sensing

The idea of using nanopores inserted in lipid membranes as a tool for the analysis of single molecules was inspired by the very intense molecular transport activity between the intracellular and the extracellular media as well as between different cellular organelles. Molecular transport across the naturally-impermeable membranous structures of cells occurs through a variety of protein channels incorporated into the fluid mosaic of the lipid bilayers. These channels or pores act as gates through which a wide variety of molecules such as ions, sugars, nucleic acids and proteins can pass during their transport from one organelle to another, or from the cytoplasm to the outside of the cell. The ability of the pores to allow the passage of ions and larger molecules suggested that the ions could be used to carry an electric current which in turn could drive larger polar molecules through the channels. The change in the ionic flow through the channel due to the transit of the macromolecule would depend on the structure of the particular translocating molecule. Similar to the natural molecular translocation mechanism through protein channels, the nanopore detection method utilizes a nanopore inserted into an insulating membrane separating two chambers filled with a buffer/electrolyte solution.

The principle for nanopore detection is similar to the Coulter counter used for counting and sizing particles [5]. An electric potential is applied across the membrane via two electrodes and the ionic current through the open pore is monitored. When a charged molecule is driven into and through the nanopore by the electric potential, it causes a drop in the ionic current. The ionic current drop has a characteristic amplitude (I_{block}) and duration (T_{block}) reflecting the particular structure of that molecule (Fig. 1.1). The amplitudes are related to the volume of the molecule while the durations depend mainly on its length and charge; the changes in the volume and length of the molecule are connected to the particular structure adopted in solution. Thus, in principle, nanopore detection could be used to distinguish between identical molecules adopting different conformations. This feature could prove useful for the investigation of protein mis-folding diseases.

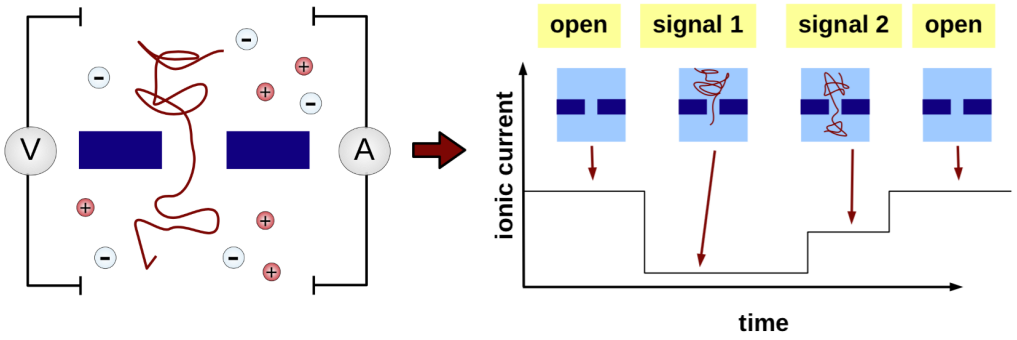


Figure 1.1: *Current through the nanopore device attracts the target in the pore. The nanometer pore allows the analysis to single molecule. During translocation the output signal changes allowing to analyze the target.*

Currently, there are two types of pores used for the nanopore detection method: protein pores and solid-state pores. The protein pores belong to the group of pore-forming toxins produced by bacteria with damaging effects on the cytoplasmic phospholipid bilayer of human and animal cells. Their innate property of auto-insertion into lipid bilayers played a crucial part in establishing this group of proteins as sensors. Solid-state pores have been developed with the goal of improving the life span of the nanopore setup by using synthetic membranes, the range of molecules that can be analyzed by controlling the pore diameter as well as the range of experimental conditions that can be used (pH, temperature, ionic strength, applied potentials, etc.).

1.1.1 Biological Nanopores

The biological nanopores possess numerous advantages which make them good targets for experimental and computational research. Firstly, their atomistic structures are usually known from X-ray crystallographic studies. Secondly, their properties may be altered through the relatively simple means of site-directed mutagenesis; it is possible not only to mutate a residue, but also to alter its properties further by coupling an additional molecule to the pore via. The heterogeneity observed among biological nanopores in terms of size and composition; the cells can produce large numbers of biological nanopores with an atomic level of precision that can not yet be replicated by the artificially.

Alpha-Hemolysin

The biological nanopores traditionally used is the heptameric protein α-Hemolysin. α-Hemolysin is a nanopore from the bacteria that causes lysis of red blood cells; its pore is a channel-protein 10 nm long with two distinct 5 nm sections made of 3.6 nm diameter vestibule connected to a transmembrane β-Barrel that is 2.6 nm wide. The very important particular is that the size vestibule is just 1.4 nm wide, which means that single-stranded DNA can pass through the nanopore, but double-stranded DNA cannot (Fig. 1.2), suggesting the potential emergence of α-Hemolysin as a next-generation sequencing tool.

The problem that the molecules move through the nanopore at high velocities (for the DNA estimated to be 1 nucleotide per microsecond) under typical experimental conditions. These velocities mean that only a small number of ions (as few as 100) are available in the nanopore to correctly identify any given nucleotide, so the small changes in the ionic current due to the presence of different nucleotides are likely to be overwhelmed by thermodynamic fluctuations. Sequencing using α-Hemolysin has been developed through basic study and structural mutations, an approaches typical is to incorporate enzymes to regulate the transport. An enzyme coupled to a nanopore is attractive for two reasons:

- The enzyme helping to capturing electrophoretically in the nanopore the molecule in solution.
- The motion is slowed and controlled from as the enzyme processes the molecules.

A example of mutation of the protein α-Hemolysin that improves the detection abilities of the pore [6] is binding of an exonuclease onto the pore.

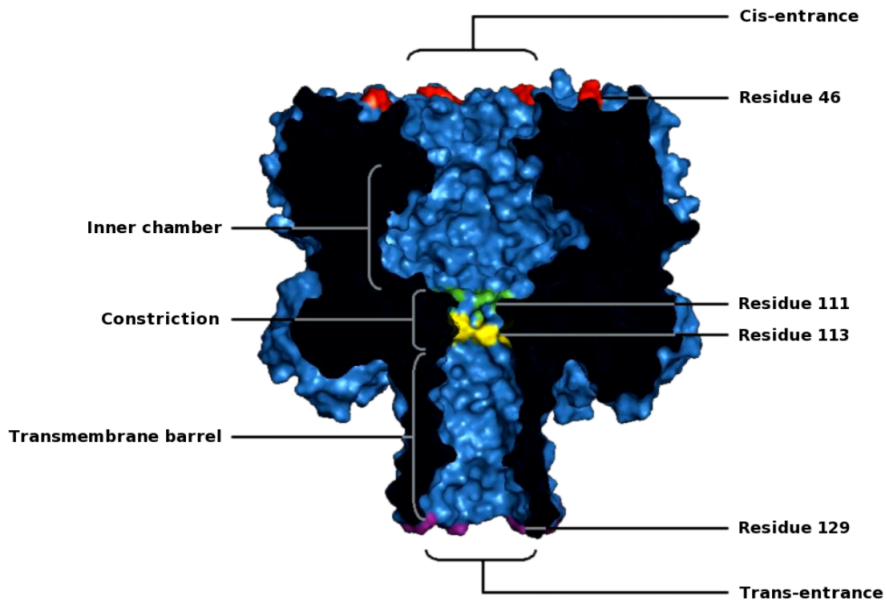


Figure 1.2: *A cross-section of the heptameric transmembrane form of α-Hemolysin. The diameters of the pore features are as follows: cis-entrance, 28 Å; inner chamber, 46 Å; constriction, 14 Å; transmembrane barrel, about 20 Å; trans-entrance, 24 Å. The height of the entire pore is about 100 Å, while the constriction to the trans-entrance measures about 52 Å. Residues of particular interest are also highlighted.*

The enzyme cleaves the single bases, enabling the pore to identify successive bases. Coupling an exonuclease to the biological pore would slow the translocation of the DNA through the pore, and increase the accuracy of data acquisition. Also recents studies have pointed to the ability of α-Hemolysin to detect nucleotides at two separate sites in the lower half of the pore [7]. The R1 and R2 sites (Fig. 1.3) enable each base to be monitored twice as it moves through the pore.

This method improves single reading through the nanopore by doubling the sites where the sequence is read by nanopore. Although α-Hemolysin has dominated the biological nanopore sequencing landscape so far, other more efficient biological nanopores are emerging. A structural drawback with α-Hemolysin is that the cylindrical β-Barrel can accommodate up to 10 nucleotides at a time, all significantly modulate the current into pore [8]; this dilutes the ionic signature of the single nucleotide in the 1.4 nm constriction, thus reducing the overall signal-to-noise ratio in sequencing applications; to solve this problem, it was estimated to truncate the β-Barrel up to the constriction [9].

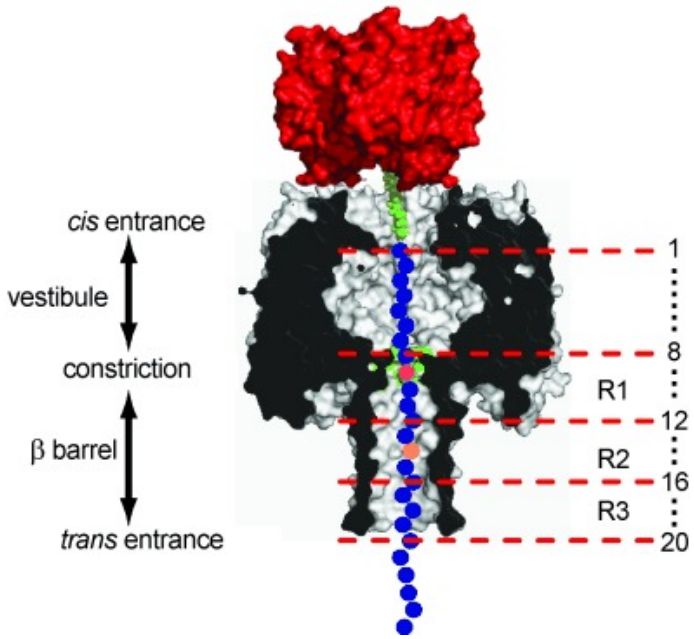


Figure 1.3: *Schematic representation of an oligonucleotide (blue) immobilized inside an α-Hemolysin pore (grey) by the use of a biotin (yellow) streptavidin (red). The α-Hemolysin pore can be divided into two halves, each approximately 5 nm in length containing a roughly spherical vestibule, the transmembrane β-Barrel, within this is located the constriction 1.4 nm diameter (green). R1, R2, and R3 represent the three base-recognition sites in the α-Hemolysin nanopore within theβ-Barrel domain of the pore.*

1.1.2 Solid-State Nanopores

Synthetic nanopores present several advantages of relevance for the biotechnological application sensing and sequencing. Firstly, they have the potential to be significantly more robust than protein nanopores, which may be destroyed or disrupted by high transmembrane potentials and extremes of pH and temperature. One can envisage such robustness being of tremendous value when applied to large scale sequencing. Not being limited to biological molecules, synthetic nanopores may be constructed from a wider variety of materials. The solid-state approach offers the ability to regulate not only the size but also the nanopore shape, with a sub-nanometre precision the ability to fabricate high-density arrays of nanopores, superior mechanical, chemical and thermal characteristics and the possibility of integrating with electronic readout techniques. This can grant them interesting properties such as electrical conductance, which, for example, allows electrical measurements to be made directly at the pore.

There has been considerable progress in engineering and harnessing synthetic nanopores over the past decade. The first sculpting of synthetic nanopores from thin solid-state membranes of Si_3N_4 with highly focused low-energy ion beams [10]; the pores fabricated with this method were asymmetrical in terms of geometry and electrical properties. Other groups [11, 12] prefer to use a electron beam with a diameter probe of few nanometres [13], these solid-state nanopore are able to selectively analize segments of single stranded DNA across the pore [14]; a major advantage of this technique is that direct visual feedback is possible through the TEM which allows controlling the pore diameter with single nanometer precision. Recently, 1-5 nm thick graphene membranes electron-beam sculpted nanopore have been developed (Fig. 1.4 a).

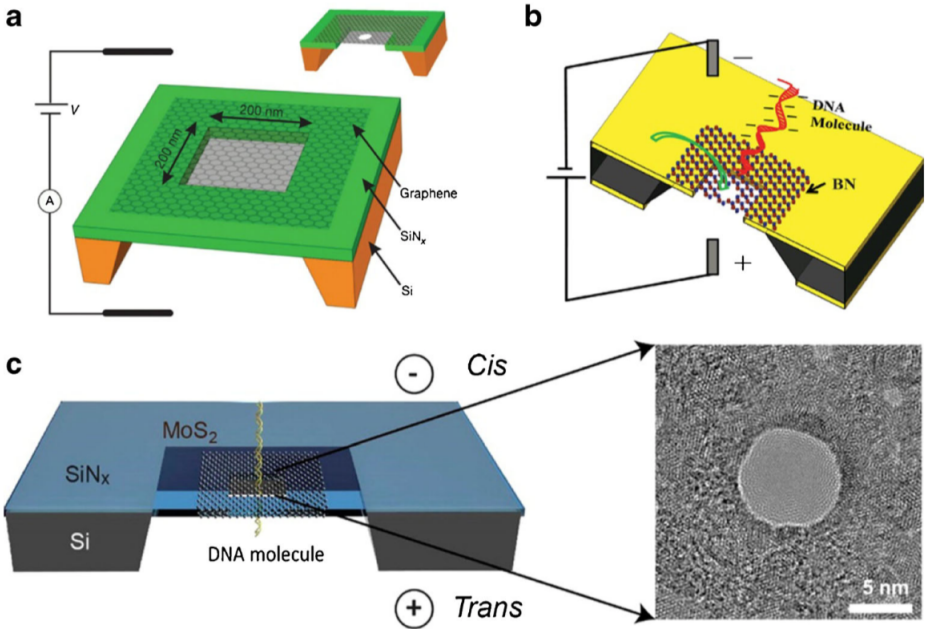


Figure 1.4: *Solid-statenanopores fabricated with two-dimensional nanomembranes: a a graphene nanopore; b a BN nanopore; c a MoS2 nanopore; [15–17].*

Graphene

Graphene is a two-dimensional sheet of carbon atoms; it possesses remarkable mechanical, electrical and thermal properties. Moreover, the thickness of a single layer of graphene is comparable to the space between nucleotides inside the DNA (0.32-0.52 nm), making it particularly attractive for single-molecule sequencing [18]. In 2008 have been fabricated the first nanopores in

graphene films and subsequent transmission electron microscopy (TEM) the following studies elucidated the kinetics of pore formation, stability of the pore in graphene and how to detect single DNA molecules using nanopores in graphene films; the films are prepared either by chemical vapor deposition or exfoliation from graphite and the nanopores (diameters 2-25 nm) were produced by a focused electron beam [19]. It was found for graphene that the conductance of the nanopore was proportional to the pore diameter, whereas the conductance is typically proportional to the square of the diameter for Si_xN_y but it is much thicker [15]. They also found that the conductance for a given diameter of nanopore remained largely constant as the number of graphene layers in the membrane was increased from one to eight. So it is possible to produce a multilayer graphene nanopore, with the number of layers increasing as we move away from the pore (Fig. 1.5).

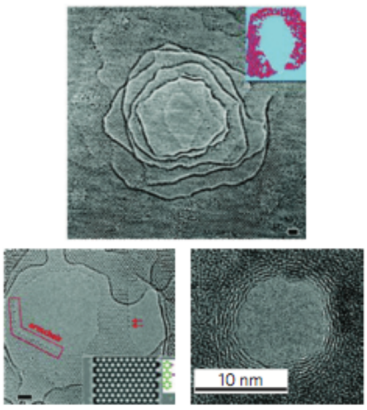


Figure 1.5: *TEM image of a terraced nanopore formed in a graphene film containing 10 monolayers of carbon atoms. Scale bar, 1 nm. Bottom left: nanopore in a monolayer of graphene with primarily armchair edges surrounded by multilayered regions. Scale bar, 1 nm. Bottom right: TEM image of a nanopore in multilayer graphene; ripples at the pore edge again show the terraced structure.*

This conformation "terraced" was confirmed using TEM image analysis [20]; a terraced nanopore architecture could prove very useful for many reasons, kind multilayered support may increase the stability and longevity of a graphene nanopore sensor. The terraced effect gives induced greater current blockades than the nanopores in single-layer graphene under normal conditions. The first computer simulations with the DNA passing through a graphene nanopore revealed a resolution similar to the size of an individual nucleotide. This result suggests that should be possible single-nucleotide detection with an electronic read.

1.2 Nanopore for Bio-Molecules

Knowledge of the bio-molecule sequence represents an opportunity onto which a broad range of biological phenomena can be studied. Over the past years, more devices for DNA sequencing have become widely available, reducing the cost, and the possibility to put the sequencing capacity of a major genome center in the hands of individual investigators. These new technologies are rapidly evolving, and near-term challenges include the development of robust protocols for generating sequencing libraries, building effective new approaches to data analysis and new experimental designs. Next-generation bio-molecules sequencing has the potential to dramatically accelerate biological and biomedical research, by enabling the comprehensive analysis of genomes. Nanopores are seen as a potential next-generation sequencer that could provide cheap and fast sequencing. A crucial problem to the success of nanopores as a reliable analysis tool is the fast and stochastic nature of the bio-molecules translocation. It searches with studies experimental new modifications for slowing and controlling the translocation, such has the incorporation of biological motors. The most studied biomolecules with nanopore systems are polynucleotides and proteins.

Polynucleotides

Polynucleotides in nature exist in two classes: deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). DNA stores genetic information for the development and function of living organisms and some viruses. RNA is transcribed from DNA by enzymes in the production of proteins, and most viruses keep their genetic information in the form of RNA. The molecular structure of DNA is shown on the left of *Figure* 1.6. Each nucleotide monomer in the polymeric chain is composed of a nucleotide base, a pentose sugar (deoxyribose), and a phosphate group [21]. The deoxyribose sugar and phosphate group form the repeat unit of the polynucleotide backbone, with each monomer connecting to the one via a phosphodiester bond. At physiological pH, the phosphate groups in the polynucleotide chain possess a single negative charge while terminal phosphate groups posses a double negative charge. The terminus that ends with a phosphate group is called the 5'-end due to the relation to the pentose sugar. The opposite end, which is terminated at a pentose sugar, is termed the 3'-end. Attached to the pentose sugar is one of four types of nucleotide base: cytosine, guanine, adenine and thymine. These nucleotide bases can form hydrogen bonds to complementary bases. This interaction is specific due to the matching molecular structures and hydrogen bonding pattern between guanine and cytosine, or adenine and

thymine. DNA is usually in the double helical form as illustrated on the right of *Figure* 1.6. In this form, one polynucleotide strand is bound to a second strand of complementary sequence [21]. The shape of the double helix reflects the conformational preferences of the nucleotides as well as the intramolecular interactions between nucleotides within one single strand and intermolecular interactions between the two complementary strands. A dominating factor which maintains the intermolecular interaction is the energetically favorable base stacking between adjacent bases. Base stacking is determined by three factors: attractive London dispersion forces, short-range exchange repulsion (which is reduced by increased π-orbital overlap), and electrostatic interactions [22]. The sequence of nucleotide bases along the polynucleotide strand encodes the genetic information to form polypeptides. In the presence of the necessary conditions and molecular machinery, DNA is transcribed into RNA, which in turn serves as a template for the biosynthesis of polypeptides [22]. Similar to DNA, each RNA nucleotide is made up of a pentose sugar, a phosphate, and a base. Unlike DNA, the pentose sugar of RNA is ribose and contains a hydroxyl group at the 2'-carbon of the pentose. Furthermore, the uracil base is used in place of thymine. RNA does not typically exist as a duplex, as found in DNA. However, single stranded RNA can form secondary structure elements such as hairpin loops, stems, bulges, and internal loops, which are largely mediated via hydrogen bond base pairing. In single stranded sections of RNA, the biopolymer's conformation is influenced by the energetically favorable overlap of the aromatic π-orbitals on the nucleotide bases. As this interaction is considerably weaker than the hydrogen bonds within a double-stranded structure, single stranded polynucleotides do not tend to form highly regular configurations. Nevertheless, the base-stacking area between adenines is one of the reasons why polynucleotides composed purely of adenosine adopt a helical shape; the hydroxyl of the ribose unit also contributes to the structure. By contrast, the smaller cytosine base in polydeoxycytidine tends to adopt a more random configuration.

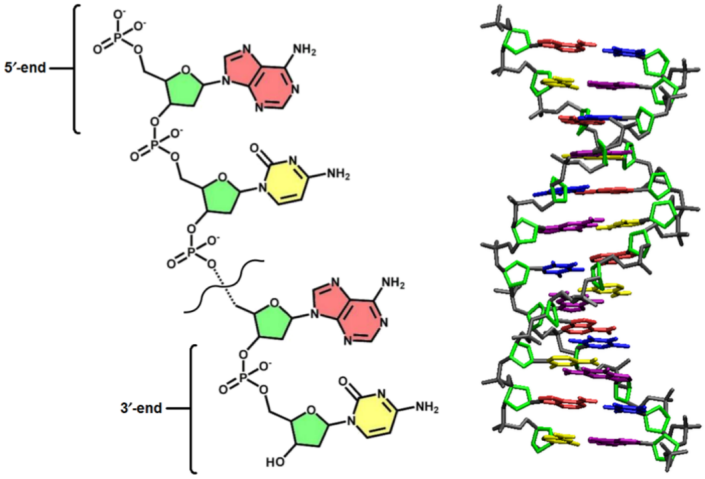


Figure 1.6: *Illustration of the molecular structure and conformation of DNA. The molecular structure of a single strand of DNA is shown on the left, featuring the phosphate groups, the deoxyribose sugar (green), adenine nucleotide bases (light red) and the cytosine nucleotide bases (yellow). The 3' and 5'-ends of the polynucleotides are also labelled for reference. The generic double stranded DNA helix conformation is shown on the right.*

Proteins

Proteins are folded biopolymers composed of interlinked amino acids. Proteins are key to the structure and composition of all living organisms and participate in almost every biological function. Amino acids possess amine and carboxyl functional groups, which form the amide bond between separate amino acids of the polypeptide chain. Amino acids also contain a side chain which varies between different amino acids types, side chains for lysine and methionine are shown as examples in *Figure* 1.7. Twenty different amino acids occur in nature, each pertaining different properties depending on the type of side chain attached. Relevant to the work in this thesis, the side chains can be neutral or charged depending on the pH of solution that the amino acid is in. The linear sequence of amino acids in a polypeptide chain represents the primary structure of proteins [23]. The polypeptide chain can form secondary structures such as α-helices (arising from hydrogen bonds between the amine and carbonyl group of nearby amide bonds) and β-sheets (sections of polypeptide connected laterally by more than 4 hydrogen bonds). The polypeptide, secondary structure included, can fold in itself forming the tertiary structure of the protein. Several proteins may bind together to form a quaternary structure. For instance, the lipid bilayer membrane protein α-hemolysin is formed by seven polypeptide chains within a heptameric struc-

ture. Protein structures can be determined using X-ray crystallography or nuclear magnetic resonance spectroscopy among other methods [24]. By obtaining the atomic coordinates for the atoms in a protein, the protein is then able to be used in computer simulations.

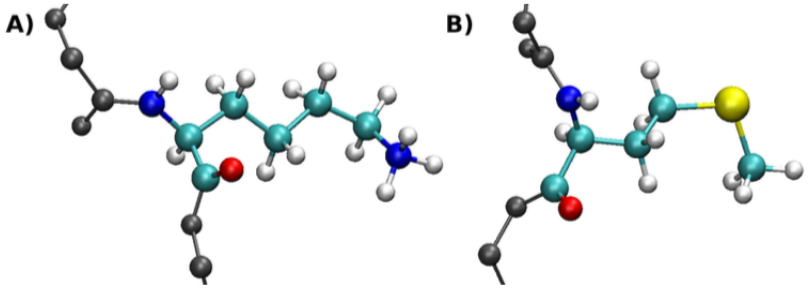


Figure 1.7: *Representation of the molecular structure of amino acids lysine and methionine. Amino acids are composed of amine and carboxyl functional groups along a polypeptide backbone (grey). They also posses side-chains, the composition of which varies between types of amino acid. An amino acid contains carbon (cyan), nitrogen (blue), and hydrogen (white), and may contain sulphur (yellow). A) Lysine is a positively charged amino acid at physiological pH, due to the terminal amino group of the side chain being protonated. B) Methionine amino acid is a neutral amino acid at physiological pH.*

1.2.1 DNA Sequencing

First-generation sequencing is based on the Sanger method, which was presented in 1977 by the scientist Frederick Sanger [1]. In this method, natural and chain-terminating nucleotides are incorporated by a polymerase into a growing DNA chain during replication. The random incorporation of chain-terminating nucleotides, which are either fluorescently or radioactively labeled during the polymerase chain reaction (PCR), leads to a population of DNA strands with different lengths. These DNA strands are then separated according their size by capillary electrophoresis. A laser combined with a fluorescence detector detects the fluorescently labeled terminated DNA when the molecules pass through the capillary, which allows them to be sequenced.

The second-generation of DNA sequencing instruments works by detecting the incorporation of the labeled nucleotides directly and prevents the necessity of separating the DNA in a gel [25]. However, since earlier optical sensors were not able to detect the incorporation of a single nucleotide, a PCR step is still needed to amplify the DNA molecules. This creates a large number of fluorescently labeled DNA molecules to generate enough photons

to excite the optical detectors. Introduction of the second-generation devices in 2007 considerably accelerated the reduction in the cost-to-base ratio (Fig. 1.8). In 2008, a second-generation device managed to sequence an entire human genome in a few weeks. This improvement reduced the speed and costs for sequencing a genome [26].

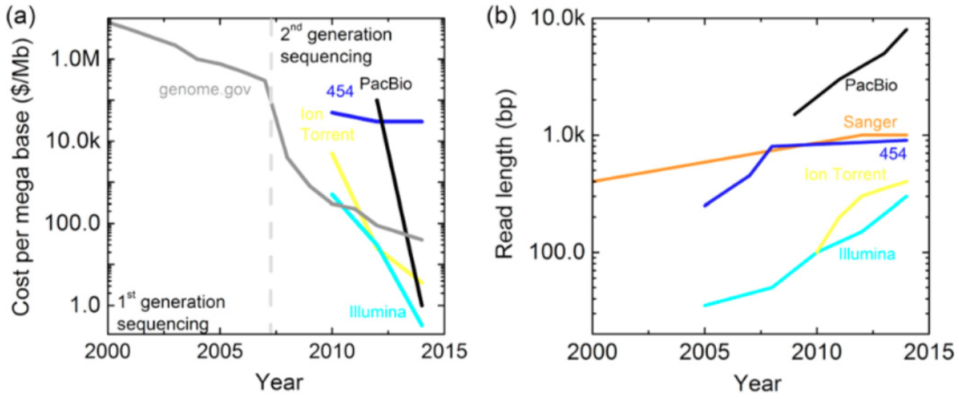


Figure 1.8: *Comparison First and Second Generation (a) Cost per base of the different sequencing techniques as a function of time. The gray curve shows data calculated by the National Human Genome Research Institute, United States, and represents the average costs, including reagents and instruments. The introduction of next-generation sequencing devices in 2007 has increased the rate of cost improvement. The blue, black, green, and magenta curves show the decline in costs for the various NGS techniques such as 454 Life Sciences, Pacific Biosystems, Ion Torrent, and Illumina, respectively. (b) Read lengths plotted against time. The orange curve represents the original Sanger sequencing technique, which was ideal for de novo genome sequencing tasks but is gradually being replaced by instruments from Pacific Biosystems. Techniques from Ion Torrent and Illumina cannot perform long read lengths but are profitable due to their low cost-to-base ratio.*

The current next-generation sequencing technologies are defined by several characteristics. First, they are able to detect a single unmodified nucleotide by relying on new optical and electrical single-molecule techniques. This avoids the need for amplification by PCR, thereby reducing time and expenses for reagents. Another advantage lies in the extended read length, which surpasses the 1000 bp limit, reaching up to 50 kbp; this increased read length also reduces the amount of reagents and time needed to sequence the DNA, thus further costs down. The nanopores are a technique of next-generation sequencing, Oxford Nanopore Technologies is developing a device based on an array of biological nanopores as described in (Fig. 1.9) and launched a test at the beginning of this year [27]. A commercial launch has not yet been disclosed, but if the technology is coupled with a device that

enables reliable decoding of long sequences with an acceptable error rate, it could change the current landscape of DNA sequencing. In particular, the low cost and footprint (weight and volume) could make these devices ideal for private users, field scientists in remote areas, and food processing industries. In order to control and reduce the translocation speed of DNA to a rate at which single-nucleotide resolution is feasible (1-100 ms/nt), the major approaches attempted are enzyme-mediated transport by incorporation of a biological motor and Voltage-driven transport controlled by adjustment of pore geometry and experimental conditions.

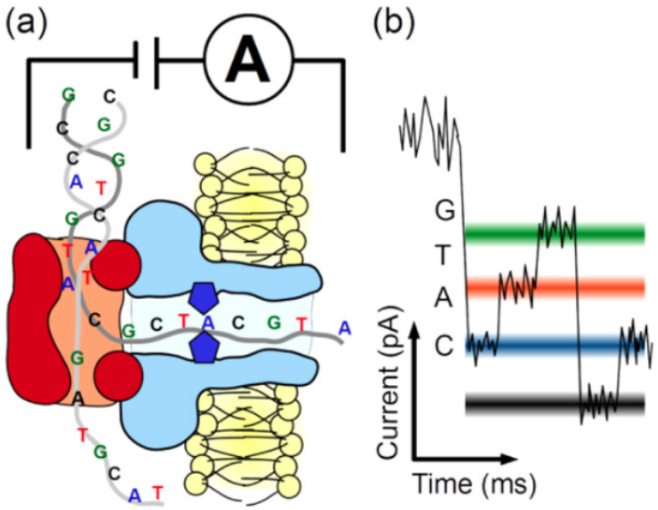


Figure 1.9: *(a) Double-stranded DNA is separated into single-stranded DNA by a polymerase such as phi29. This decelerates the translocation velocity of the ssDNA through the nanopore. The nanopore possesses a constriction inside the channel (dark blue diamond), which enables reading of the ssDNA sequence. (b) Simplified scheme depicting the decoding method. The ionic current trace is altered by the ssDNA sequence translocating through the nanopore, as shown in figure(a). Each level represents one nucleotide residing inside the nanopore at a specific point in time. By detecting these levels, the sequence of the DNA can be decoded. In the current experiment, the nanopore is not sensitive enough to detect one single nucleotide, but four nucleotides can increase the number of levels to 256.*

In conclusion, nanopores have tremendous potential for revolutionizing nucleic acid analysis, specifically DNA sequencing. There is no equivalent nanoscale device to a nanopore that allows localization/ transport of DNA sequentially in space, in addition to detect its identity. However, controlling DNA motion through a nanopore, and read-out of its sequence, are independently grand challenges that require further development before this method can be usable as device.

1.2.2 Proteins and Peptides Sequencing

In parallel with the nanopore analysis of DNA, several groups have reported that peptide and protein molecules could be analyzed in the same way. Based on the idea proposed by Singer that polypeptides translocate through channels in vivo [28]. Initial studies on peptides showed that different sequences have different values of I_{block} and T_{block}. For instance, while for the α-helical peptides, both I_{block} and T_{block} increase with the length, on the other hand for the β-sheet structures smaller values of I_{block} and T_{block} are reported compared with α-helix (Fig. 1.10). This breakthrough indicated that nanopore sensing could be used to conduct structural and conformational studies of peptides and proteins, enabling the collection of results difficult to obtain by bulk spectroscopic techniques such as circular dichroism (CD) or nuclear magnetic resonance (NMR). Thus, although it is not possible to identify with certainty a particular structure based on the values of I_{block} and T_{block}, tentative assignments can be made comparing them with previous data [29].

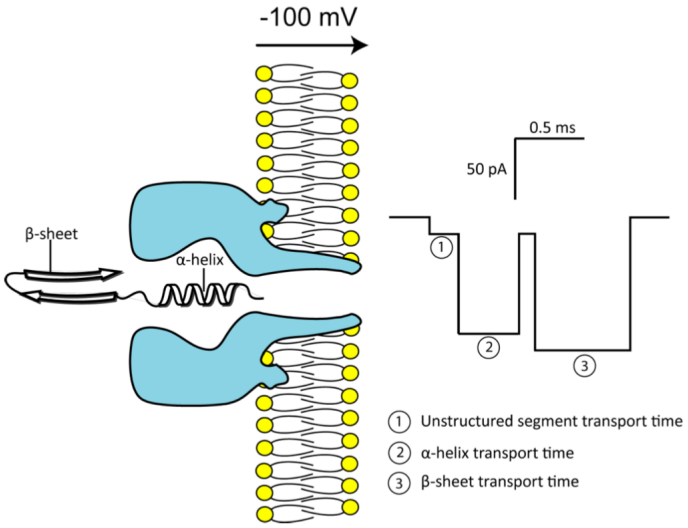


Figure 1.10: *Schematic representation of the α-hemolysin pore inserted into a lipid bilayer with α-helical and β-sheet segments of a protein ready for translocation. The transit of differently folded segments should be reflected within the translocation profile of each protein molecule.*

Furthermore, the peptides had much lower charge densities than DNA which was reflected in transit times of 1 to 2 orders of magnitude longer, resulting in increased signal resolution compared to nucleic acids. It was found that the number of events per time increased with the applied electric potential and decreased with the peptide length [30]. However, sequencing peptides

and proteins by nanopores are even more challenging because there are 20 common amino acids and some of which are chemically very similar, there is also the problem that neutral peptides cannot be driven through the pore. The richness of information obtained from the nanopore analysis of peptides proved that different secondary and tertiary structures can be individually identified and distinguished, suggesting that nanopore sensing could be used to help elucidate one of the most intriguing problems of modern biochemistry: the folding mechanism of proteins [31]. Protein misfolding is believed to be the primary cause of Alzheimer's disease (AD), Parkinson's disease, Huntington's disease, the "prion" diseases such as Creutzfeldt-Jakob disease, cystic fibrosis, amyotrophic lateral sclerosis and many other degenerative and neurodegenerative disorders. Regardless of the type, the post-mortem analysis of brain tissue shows the presence of amyloid fibrils, plaque which consist of protein aggregates. Surprisingly, the proteins show no obvious sequence or structural homology; the fact that the proteins are not in their native conformations has led to the hypothesis that they are all "protein misfolding diseases" [32]. Misfolded proteins are normally sequestered or neutralized by cellular defense mechanisms which include the chaperone, proteasome and/or auto-phagosome responses, one possibility is that these responses are affected during AD as well as the normal protein turnover, which is essential for cell survival, not for functional aspects. Whatever is the real mechanism of action, the presence, and the subsequent misfolded peptides play a central role in this pathological condition [33]. It is perhaps surprising that there is no definitive structure for any of the misfolded proteins. The problem is that at high concentrations they inevitably aggregate which precludes the use of NMR or X-ray crystallography. Nanopore analysis is ideally suited for studying peptides which can adopt multiple conformations since each molecule is interrogated individually (Fig. 1.11). It is also very sensitive since in theory a single molecule can be detected. Therefore, the experiments can be performed at relatively low concentrations so that aggregates will only form slowly if at all.

The issue in protein's measurement is the transport of the analysis to and through the pore, with electrophoresis, which is very useful in nanopore DNA measurement, is less effective for neutral or weakly charged species. A technique using unmodified nanopores that facilitates protein translocation and enables specific protein identification [34]. The use of double-stranded DNA carrier molecules capable of binding one or more analytes. The highly charged DNA can electrophoretically transport the bound protein through the nanopore for sensing, and the binding site selectivity enables highly specific detection. This approach was also able to identify a single analyte from a mixture. The Synthesis of DNA strands capable of binding to specific

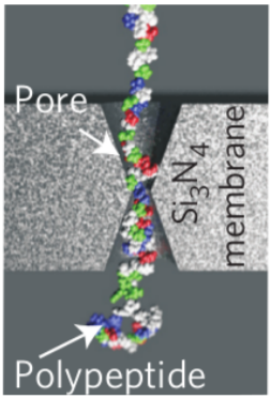
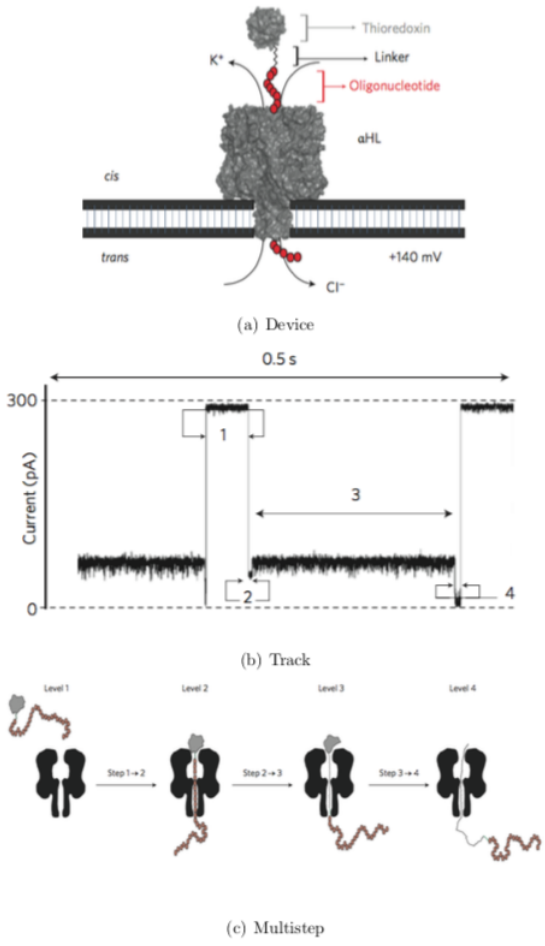


Figure 1.11: *Schematic of the translocation of a protein through a nanopore. Denaturing agents impart a uniform negative charge to the protein, resulting in a rod-like structure*

biomolecules is able to give a binary indication of their presence in a solution containing the target and a background mixture. With further improvement in data analysis or binding site design, quantitative concentration determination may also be possible. This technique is highly generalizable a DNA carrier can be easily designed and created for any combination of specific ligand-receptor binding pairs. The adaptability of this platform opens many possibilities including detection of antibodies and single molecules inaccessible with other techniques. It is demonstrated the ability of nanopores, to distinguish between folded and unfolded states of proteins at the single-molecule level. These experiments reveal additional detail, such as the existence of intermediates. In total, it observes different steps in the translocation process (Fig. 1.12). The pull causes partial unfolding, the unfolding rate has an exponential dependence on the applied potential. The partial unfolding destabilizes the remainder of the folded structure, which then unfolds spontaneously and diffuses through the pore. The experiments showed also how the passage through the pore depending on the applied potential and the exit of the pore independent of voltage [35].



(a) Device

(b) Track

(c) Multistep

Figure 1.12: *(a) The pore is inserted into a lipid bilayer from the cis compartment and a potential is applied, causing an ionic current to flow through the pore. (b) Current trace translocation. (c) Level 1: Peptide (V5-C109-oligo(dC)30) is in solution and the pore is unoccupied. Level 2: the oligonucleotide threads into the pore and pulls on the protein. Level 3: the pulling force causes partial unfolding, allowing the oligonucleotide to traverse the pore and the unfolded segment of the polypeptide to enter. Level 4: the remainder of the polypeptide unfolds spontaneously, diffuses through the pore and leaves through the trans entrance.*

In conclusion, many neurodegenerative diseases involve protein misfolding into insoluble fibrils or amyloid plaques. The misfolding processes, particularly the first steps, are very difficult to study by conventional techniques because at high concentrations the proteins inevitably aggregate. Nanopore analysis, being a single molecule technique, is an attractive alternative because it can be performed at very low concentrations.

1.3 Atomistic Approaches to Nanopore

Computer simulations of molecular models can give insight into microscopic processes such as nanopore translocation to further our understanding of experimental observations, potentially providing the basis for new or refined experimental approaches. Recent advances in genetic sequencing using nanopores implicate a fast and cheap method of sequencing, given suitable refinement. A greater understanding of the microscopic translocation factors could greatly aid in this regard, which may be achieved through simulation. Given modern resources one cannot arbitrarily replicate nature in simulations to a subatomic level of precision, however, there are techniques for replicating nature in probabilistic detail, atomic detail, even sub-atomic quantum detail [36].

Molecular Dynamics (MD) simulations calculate the Newtonian equations of motion of all particles in the system based on the forces present and recalculate the positions of the particles for a small increment in time, usually 1 or 2 femtoseconds. The deterministic time-evolution of the system is calculated in this way, the accuracy of which depends on the level of detail in the model and interactions supplied and allowed. Due to current limits in the timestep of each recalculation, there are limits on the timescales that can be simulated in atomistically detailed MD. Such limits are typically in terms of nanoseconds, but with the increasing availability and scope of high-performance computing facilities, there are instances of microsecond timescales [37] and even millisecond timescales [38]. Since MD simulations reproduce the deterministic time-evolution, they allow for the calculation of the dynamics of the system.

Simulations can also be performed at the quantum sub-atomic level. In MD simulations, the model of interactions between the atoms of the system is supplied as the simulation's input, which requires prior knowledge about the interactions of the system. *Ab-initio*, on the other hand, uses the laws of quantum mechanics to calculate atomic interactions. The main value in *Ab-initio* is that bond breaking and bond forming is accounted for, unlike in classical MD.

Chapter 2

Theoretical Methods

In this chapter methods for the computer simulations are detailed. Using these methods it is possible to gain insight into such systems and produce important features of a process. Through a greater level of understanding unlocked by simulations, we can understand more about biological processes, and it is possible to design improved experimental components and conditions.

2.1 Molecular Dynamics

In this approach all atoms in the system are treated as classical particles moving under the influence of the Newton's classical equations of motion [39].

$$m_i \frac{d^2}{dt^2} \vec{R}_1(t) = \vec{F}_i(t) = -\nabla V(\vec{R}_i(t), ..., \vec{R}_n(t)) \tag{2.1}$$

Where $\vec{R}_i$ and m_i denote the position and the mass of the atom i respectively, and n the total number of atoms in the system. The force $\vec{F}_i$ acts on atom i and is determined as the gradient of the potential energy $V(\vec{R}(t))$ of the system. The empirical force fields describe the potential energy of the system in terms of the interactions between the atoms. In its functional form, the potential energy function has two kinds of parameters, bonded and non-bonded ones. The former class consists of potentials for bond lengths, bond angles, improper dihedral angles as well as torsional angles. The later class contains the non-bonding terms, the electrostatic as well as the Van der Waals interactions. The parameters for the force fields are obtained with quantum-mechanical calculations which can be further improved by fitting to experimental data. Among the force fields commonly used in MD simulation of biomolecules there are CHARMM [40], AMBER and GROMOS. The

various force fields may differ in the functional form of individual terms and especially in the parametrization procedures for the large number of parameters involved. One of the biggest advantages of the MD technique is the brief time required to obtain the atomic details and the structural analysis of the molecular system. Unfortunately this advantage comes at a large computational cost. Moreover, one has to keep in mind that various biological processes occur at time scales ranging from femtoseconds over milliseconds to seconds (ex. protein folding being on the slower side of this spectrum). This has to be seen in connection to the usual MD time step in the femtoseconds range for state-vibrational (such as hydrogen bond vibrations). With the present computational resources, simulations in the range of a few hundreds of nanoseconds are the common practice which may be extended up to several microseconds [41].

Although MD simulations have been very successful providing microscopic details of biological processes, several limitations associated with the method have to be taken into account. Standard force fields used in biomolecular simulations account for electrostatic interactions and the corresponding interactions are often based on a simple pairwise-additive models, and real physical systems undergo substantial polarization when placed into a medium with a high dielectric constant such as water or in the presence of an external applied electric field. Polarizable force fields are available to explicitly account for many-body induced polarization effects but they are computationally more expensive [42].

2.1.1 Steered Molecular Dynamics

Steered Molecular Dynamics (SMD) provides a means of retrieving more data in a smaller timescale while keeping atomistic detail. This comes at the cost of the system being at non-equilibrium. Approximations do not necessarily have to take the form of model simplifications, MD programs such as NAMD [43] allow simulated processes to be steered by introducing non-equilibrium forces. The SMD allows to apply a directional force an atom, causing the atoms and anything coupled to the atoms to move along the direction of the force. The benefit of this is that a significant degree of movement can be induced in a relatively short timeframe, and the full atomistic dynamics of the process can be examined. Applying SMD to threading molecule's chains through nanopores, the translocation process can be replicated in a atomistic model within a timeframe that can be simulated. By pulling the molecule strand at constant velocity (known as constant velocity steered MD or cv-SMD [43]), a translocation of known distance can be performed. In constant velocity SMD, an atom or the centre of a group of atoms is harmonically

restrained to a point in space that is shifted in the chosen direction. The harmonic restraint can be thought of as a spring attached to a dummy atom, the strength of the restraint is given by a force constant k (figure 2.1). The simulation outputs the force in pico-newtons (pN) experienced by the spring in the direction of pulling (the reaction coordinate). This dummy atom is moved at constant velocity and then the force between both is measured using:

$$\vec{F} = -\nabla U \tag{2.2}$$

$$U = \frac{1}{2}k[vt - (\vec{r} - \vec{r}_0) \cdot \vec{n}]^2 \tag{2.3}$$

Where U is potential energy, k is spring constant, v is pulling velocity, t is time, $\vec{r}$ and $\vec{r}_0$ are actual position and initial position of the SMD atom, $\vec{n}$ is direction of pulling.

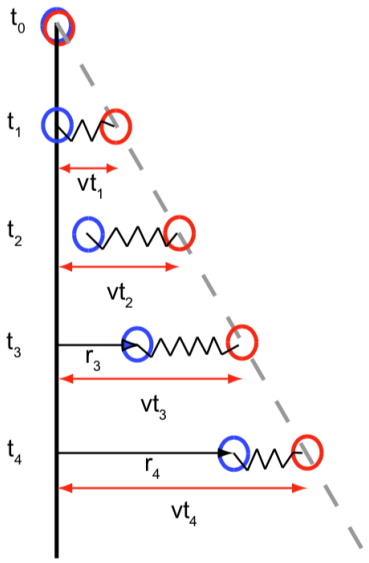


Figure 2.1: *Pulling in a one-dimensional case. The dummy atom is colored red, and the SMD atom blue. As the dummy atom moves at constant velocity the SMD atom experiences a force that depends linearly on the distance between both atoms.*

An alternative to cv-SMD is constant force SMD (cf-SMD), here the constraint point is moved to keep the force on the restraint at a constant value, resulting in a variable velocity. This is advantageous when the applied force must remain limited, but one loses control over translocation time, hence the length of the simulation. Therefore cf-SMD is not applicable for these simulations.

Ionic Current

Ionic current measurements are used to characterize single nanopores and their interactions with biological molecules; the current is calculated with the equation 2.4

$$I(t + \Delta t/2) = \frac{1}{\Delta t l_z} \sum_{i=1}^{N} q_i(z_i(t + \Delta t) - z_i(t)) \tag{2.4}$$

Where z_i and q_i are respectively the z-coordinate and charge of ion i, Δt is the simulation time and l_z is the lenght of the cell along z.

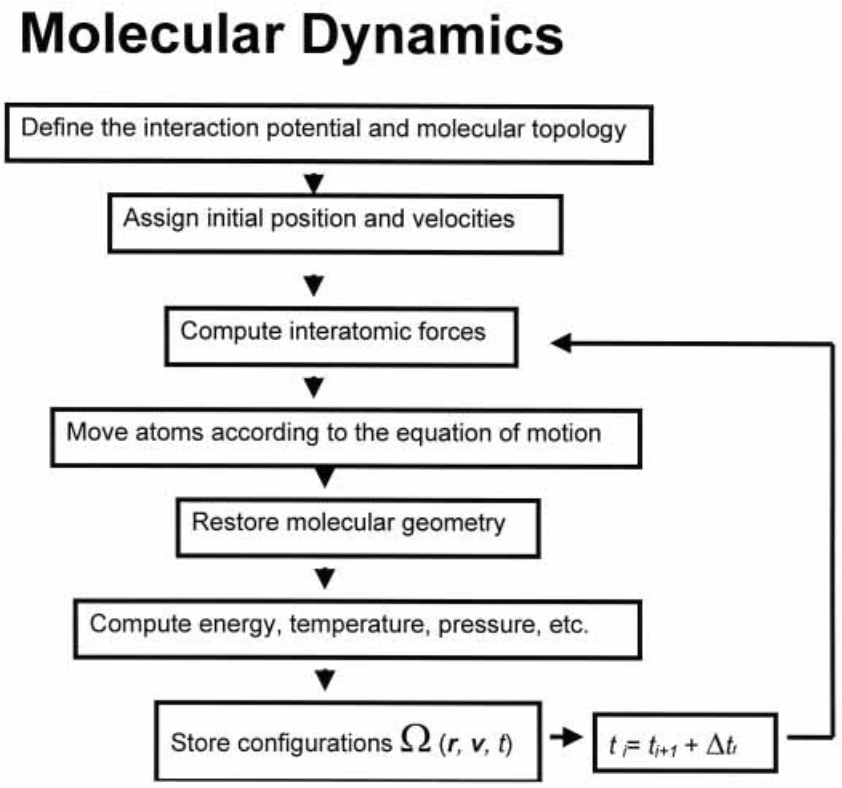


Figure 2.2: *Scheme of the molecular dynamics simulation procedure.*

2.2 Density Functional Theory

Density Functional Theory (DFT) is a method which determines the ground state of a system of N electrons. The DFT is one of the most popular ab initio methods which is interested in the total electronic density at each point in space, rather than attempting to obtain the many-particle wavefunction directly. DFT has proved to be succesful in accounting for structural and electronic properties of a vast class of materials, ranging from atoms and molecules to crystals and other complex extended systems.

The fundamental Schrödinger equation governing the full quantum mechanics of atoms and electrons has been well known since the 1920's [44]. However, the task of solving these equations for realistic systems is tremendous. For electrons and nuclei interacting through the Coulomb force the Hamiltonian is:

$$\begin{aligned}\hat{H} &= \hat{T}_e + \hat{V}_{e-e} + \hat{V}_{e-n} + \hat{T}_n + \hat{V}_{n-n} \\ &= -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 + \sum_{i\neq j}\frac{e^2}{2|\vec{r}_i - \vec{r}_j|} - \sum_{i,A}\frac{Z_A e^2}{|\vec{r}_i - \vec{R}_A|} \\ &\quad - \sum_i \frac{\hbar^2}{2M_A}\nabla_A^2 + \sum_{A\neq B}\frac{Z_A Z_B e^2}{2|\vec{R}_A - \vec{R}_B|} \qquad (2.5)\end{aligned}$$

The first two terms represent the electronic kinetic and Coulomb electrostatic potential energy, the third is the electron-ion attractive potential energy while the last two terms are the ion kinetic and repulsive potential energy respectively; $\vec{r}_i$, m_e and $-e$ the position, mass and charge of the i'th electron, and A labels the corresponding nucleus parameters. Within the Born-Oppenheimer approximation [45] one decouples the electrons and nuclei by neglecting the nuclei kinetic energy. Due to the large mass difference ($m_A/m_i \approx 1823Z_A$) the nuclei move on a much longer time scale and the electronic structure only depends parametrically on the nuclear coordinates. Hence one thinks of the Coulomb repulsion between electrons and nuclei together with external fields as an external potential. The solid success of DFT lies in the possibility of a complete description of a many-body ground state from the density of an noninteracting system with an effective potential. This is the essence of the Hohenberg-Kohn theorems [46] applied in the KS-theory due to Kohn and Sham [47]. According to the theorem by Hohenberg and Kohn a system's external potential and the corresponding ground state energy is a unique functional of the ground state density. Therefore, the full Hamiltonian and all properties can be regarded

as a functional of the ground state density. In the Kohn-Sham procedure one replaces the Hamiltonian of the interacting many-body problem with that of a noninteracting auxiliary system with an effective potential reproducing the correct ground state energy. In this way one reduce the problem to having to solve a sequence of single-particle equations self-consistently with a generally unknown exchange-correlation potential describing all quantum many-body interactions. A variety of different exchange-correlation functionals exist. In this project only the most common Local Density Approximation (LDA) or Generalized Gradient Approximation (GGA-PBE) are used for exchange and correlations [48].

The Hohenberg-Kohn Theorems

At the base of the DFT as it is known today, there are Hohenberg-Kohn Theorems published in 1964 [46]. The first Hohenberg-Kohn theorem states that the electron density is uniquely determined by the Hamiltonian operator and thus so are all the properties of the system. This implies that the external potential $V_{ext}(\vec{r})$ is (to within a constant) a unique functional of electronic density $\rho(\vec{r})$. As the external potential $V_{ext}(\vec{r})$ specifies the whole Hamiltonian, the full many particle ground state is a unique functional of the electron density $\rho(\vec{r})$. Thus, $\rho(\vec{r})$ determines the number of electrons N and the external potential $V_{ext}(\vec{r})$ and consequently all the properties of the ground state. The total energy functional $E[\rho]$ can be written as,

$$E[\rho] = \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r} + F_{HK}[\rho] \tag{2.6}$$

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho] \tag{2.7}$$

The first term in Equation 2.6 represents the potential energy due to the electron-nuclei interaction, and $F_{HK}[\rho]$ is an unknown, but otherwise universal functional of the electron density $\rho(\vec{r})$ only. If it was known, the solubility of Schrödinger equation would be possible for any system. The $F_{HK}[\rho]$ functional consists of a kinetic energy functional $T[\rho]$ and electron-electron repulsive interaction functional $E_{ee}[\rho]$. The second Hohenberg-Kohn theorem is nothing more than the variational principle. It states that the electronic density that minimizes the total energy is the exact ground state density. That is for a trial density $\tilde{\rho}(\vec{r})$ we have

$$E_0 \leq E[\tilde{\rho}(\vec{r})] = T[\tilde{\rho}] + \int \tilde{\rho}(\vec{r}) V_{ext}(\vec{r}) d\vec{r} + E_{ee}[\rho] \tag{2.8}$$

which means that the energy resulting from Equation 2.6, using the trial density $\tilde{\rho}(\vec{r})$, represents an upper bound to the true ground state energy E_0. E_0 is the outcome if and only if the exact ground state density is used in Equation 2.8.

The Kohn-Sham Theory

The ground state energy of a system is by given the minimum of the Equation 2.8.The functional $F_{HK}[\rho]$ can be defined as

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho] = T[\rho] + J[\rho] + E_{nc}[\rho] \tag{2.9}$$

where $T[\rho]$, $J[\rho]$, and $E_{nc}[\rho]$ are respectively the kinetic energy, the classical Coulomb interaction and the non-classical functional; of these functionals, only $J[\rho]$ is known. The Kohn-Sham theorem provides an effective way to express $T[\rho]$ and $E_{nc}[\rho]$ [47]. The idea of the Kohn-Sham approach is to use the kinetic energy of a fictitious non-interacting system T_s with the same electronic density as the real interacting one instead of the true kinetic energy $T[\rho]$, where

$$T_s = -\frac{\hbar^2}{2m}\sum_{i}^{N}\langle\psi_i|\nabla^2|\psi_i\rangle \tag{2.10}$$

$$\rho_s(\vec{r}) = \sum_{i}^{N}\sum_{\varsigma}|\psi_i(\vec{r},\varsigma)|^2 = \rho(\vec{r}) \tag{2.11}$$

$\psi_i(\vec{r},\varsigma)$ are then the spin-orbital wave functions of the fictitious non-interacting system. The two kinetic energies of the real and fictitious system are different, however, this difference is taken into account by introducing a new functional called exchange-correlation energy which contains everything that is unknown for us. Hence the functional $F_{HK}[\rho]$ takes the following new form

$$F_{HK}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.12}$$

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + (E_{ee}[\rho] - J[\rho]) = T[\rho] - T_s[\rho] + E_{nc}[\rho] \tag{2.13}$$

Calculating the $T_s[\rho]$ requires knowing the orbitals ψ_i of the non-interacting system. If the potential is characterized by the same density as the real system, it can be re-written Equation 2.6 using Equation 2.12

$$E[\rho] = \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r} + T_s[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.14}$$

Using Equations 2.9 through 2.12 in 2.14 gives the set of Kohn-Sham equations:

$$(-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}(\vec{r}_1))\psi_i = \epsilon_i \psi_i \tag{2.15}$$

$$V_{eff}(\vec{r}_1) = \int \frac{\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_2 + V_x(\vec{r}_1) - \sum_A^M \frac{Z_A e^2}{|\vec{r}_1 - \vec{r}_A|} \tag{2.16}$$

The second term in effective potential V_{eff} represents the exchange-correlation potential and defined as the derivative of energy functional with respect to electronic density

$$V_{xc} = \frac{\partial E_{xc}}{\partial \rho} \tag{2.17}$$

Knowing the three terms in Equation 2.16, determines the potential V_{eff} which consequently determines the energy and electronic density of the ground state. As the potential V_{eff} depends on density, the Kohn-Sham equations have to be solved iteratively. It is worth pointing out that if the exact forms of E_{xc} and V_{xc} were known, the Kohn-Sham approach would lead to the exact ground state energy. These functionals have unknown forms and using the Kohn-Sham in practice requires finding reasonable approximations.

The Exchange-Correlation Functionals

The local density approximation LDA

The first expression for the exchange-correlation functional E_{xc} was originally derived to describe the exchange effects in a homogeneous electron gas (HEG). In HEG, the electrons move in the presence of a background of positive charge which ensures the overall charge neutrality of the system. As the exchange energy of a HEG depends upon the local value of electronic density, the term "local density approximation (LDA)" is used to refer to those functionals based on HEG. The LDA is the basis of all approximate exchange-correlation functionals. According to LDA, the energy functional E_{xc} can be expressed in the form

$$E_{xc}^{LDA}[\rho] = \int \rho(\vec{r}) \epsilon_{xc}([\rho], \vec{r}) d\vec{r} \tag{2.18}$$

where $\epsilon_{xc}([\rho], \vec{r})$ is the exchange-correlation energy per electron, at position $(\vec{r})$, of a uniform electron gas of density $\rho(\vec{r})$. For the unpolarized system, this energy depends only on the value of electronic density in some neighborhood of position $(\vec{r})$. In the LDA approximation the exchange potential is equal to

$$V_{xc}^{LDA}(\vec{r}_1) = -\frac{e^2}{\pi}[3\pi^2\rho(\vec{r}_2)]^{1/3} \tag{2.19}$$

The generalized gradient approximation GGA

In order to improve the LDA, the so-called Generalized Gradient Approximation (GGA), where the gradient of the density is also considered, has been introduced [49]. In comparison with LDA, GGA tends to improve total energies, atomization energies, energy barriers and structural energy differences. In GGA, the exchange-correlation energy depends both on the homogeneous electron gas density and on its gradient:

$$E_{xc}^{xc}[\rho] = \int f(\rho(\vec{r}), \nabla\rho(\vec{r})d\vec{r} \tag{2.20}$$

where f is a parametrized analytic function. To obtain reasonable results the function f must be chosen with care, because the expression 2.20 does not derive from a physical system. The xc-functional used for most of the calculations presented in this thesis is the Perdew-Burke-Ernzerhof (PBE) functional [48].

Pseudopotentials

The pseudopotentials replace the strong Coulomb potential of the nucleus and the effects of the tightly bound core electrons with an ionic potential acting on the valence electrons. The valence electrons are the principal responsible for chemical bonds in materials. In some cases the wavefunctions representing valence electrons of some external shell (3d or 4f for instance), strongly oscillates, and so for sampling these wavefunction a huge number of plane waves is needed, wasting the computational efficiency of the simulation. Thus, the main reasons for the introduction of pseudo-potentials are reduction of basis set size, reduction of number of electrons and inclusion of relativistic or other effects. Electron-ion core interactions are typically represented by a non-local norm-conserving pseudopotential (NCPP), a soft potential for valence electrons only, so that core electrons disappear from the calculation; for a given reference configuration they must meet some reasonable conditions:

- the electron energy level must be the same in full all-electron calculations and calculations made with pseudopotentials $\epsilon^{ps} = \epsilon^{ae}$
- the new wavefuction obtained via pseudopotentials must be nodeless.
- the two wavefunctions must coincide from a certain value of the radial variables in order the reproduce the scattering properties of the true potential $\phi_l^{ps}(r) = \phi_l^{ae}(r)$ or $r > r_c$.
- the charge inside r_c for each wavefunctions agrees (norm-conservation):

$$\int_{r<r_c} r^2 dr |\phi_l^{ps}(r)|^2 = \int_{r<r_c} r^2 dr |\phi_l^{ps}(r)|^2 \tag{2.21}$$

where $\phi_l(r)$ is the radial part of the atomic valence wavefunction with angular momentum specified by the quantum number l and the core radius r_c is approximately at the outermost maximum of the wavefunction.

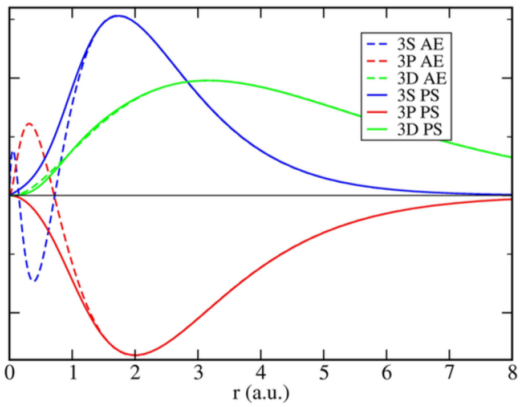


Figure 2.3: *All electron calculations vs. Pseudopotential for the radial part of wavefunctions 3s, 3p and 3d.*

2.3 Quantum Transport Theory

According to Ohm's law, the resistance R of a conductor is directly proportional to its length L and inversely proportional to its cross-sectional area A; that is, $R = V/I = (\rho L)/A$, where ρ is the resistivity and it is characteristic of the conductor material. The reciprocal of the resistance is the conductance G, $G = 1/R = \sigma A/L$, where $\sigma = 1/\rho$ is the conductivity. Equation of G suggests that the conductance of a conductor increases indefinitely by decreasing its length. However, the experimental results showed that this ohmic behaviour breaks down at some critical value of the conductor length. Below this specific length, the conductance reaches its maximum value and decreasing the length has no effect on G. In other words, as illustrated schematically in Figure 2.4, the resistance of any conductor cannot be reduced to less than a minimum finite value R_0.

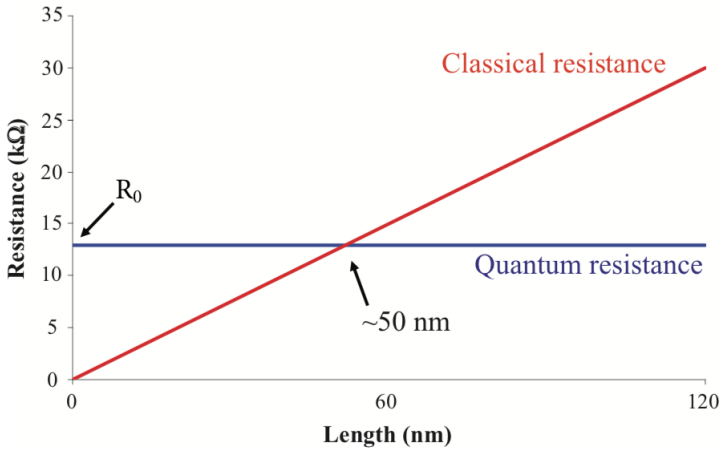


Figure 2.4: *The resistance of a conductor vs its length. Adapted from "www.quantumwise.com"*

Thus, concepts like Ohm's law are not applicable at the atomic scale. As the length of the conductor is reduced to the atomic scale such that the conductor is just an individual molecule, the effect of quantum phenomena becomes more important and dominates the conducting process; the finite resistance is associated with the resistance arising at the interfaces between leads and the sample in between them [50]. This can lead to an unexpected and incongruous behaviour to the classical one, such as a resistance that is independent of molecule length, non-linear $I - V$ characteristics and even negative differential resistance. Atomic-size conductors are a limiting case of systems in which quantum coherence plays a central role in the transport properties. A conductor would exhibit various transport regimes depend-

ing upon its dimensions relative to three characteristic length scales [51]; a fundamental length scale is the phase coherence length, L_ϕ which measures the distance over which the phase of the electron wave function is preserved (phase coherence can be destroyed by inelastic scattering mechanisms like electron-electron or electron-phonon interactions), the mean free path l which measures the distance between static collision with static scatterers and the de Broglie wavelength λ [52]. These regimes can be summarized as the following:

- Mesoscopic if $\lambda \ll L < L_\phi$
- Diffusive if $L \gg \lambda$ and $L \gg l$
- Ballistic if $\lambda \gg L > l$
- Quantum transport if $\lambda \sim L$

The diffusive and ballistic transport can be described using classical or semi-classical laws, in a semiclassical model the electron motion is a random walk of steps l among the impurities; when we reach the ballistic regime the electron momentum can be assumed to be constant and only limited by the scattering with the geometric boundaries of the sample. Whereas quantum transport phenomena (where the typical dimensions of the sample are within atomic-scale) are fully governed by quantum mechanics. The mesoscopic regime would be a separate kingdom governed by separate laws that are neither purely quantum nor purely classical; rather, a synthesis of the two. If the dimensions of the conductor are much larger than each of the three length scales, the conductor shows ohmic behaviour.

2.3.1 Elastic Transport

To study the quantum transport through an individual molecule, we will introduce a simple model for electronic transport [53]; Figure 2.5 shows the energy level diagram of a molecule with a single energy level sandwiched between metal leads.

Both leads are assumed to contain a continuum of states, which extends to all energies, and which is filled up to some Fermi level μ_L and μ_R; the molecule is coupled to both leads with coupling constants Γ_L and Γ_R which have units of energy. Physically, Γ_L and Γ_R (divided by $\hbar$) represent the rate at which an electron in the molecule's energy level ϵ would escape into left lead and right lead, respectively. The stronger coupling the higher probability

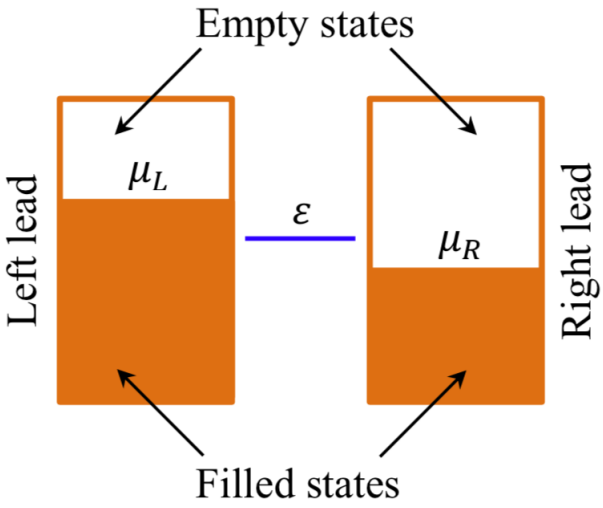


Figure 2.5: *A molecule with one energy level in between two leads.*

for escaping. If the energy level of the molecule was in equilibrium with the left lead, then the number of electrons occupying the level would be given by

$$N_L = 2f(\epsilon, \mu_L) \tag{2.22}$$

$$f(\epsilon, \mu) = \frac{1}{1 + exp\frac{\epsilon - \mu}{k_B T}} \tag{2.23}$$

where the factor 2 is for spin degeneracy and f is the Fermi distribution function. Similarly, if the level was in equilibrium with the right lead, the number of electrons occupying the level would be, $N_R = 2f(\epsilon, \mu_R)$. Under non-equilibrium conditions, the number of electrons N will be somewhere in between N_L and N_R. To determine this number, the current is written I_L at the left junction and I_R at the right junction:

$$I_L = \frac{e\Gamma_L}{\hbar}(N_L - N) \tag{2.24}$$

$$I_R = \frac{e\Gamma_R}{\hbar}(N - N_R) \tag{2.25}$$

So that from equations of I_L, I_R is obtained the current:

$$I = \frac{2e}{\hbar}\frac{\Gamma_L\Gamma_R}{\Gamma_L + \Gamma_R}[f(\epsilon, \mu_L) - f(\epsilon, \mu_R)] \tag{2.26}$$

Equation 2.26 illustrates few basic facts about the process of current flow:

- No current will flow if $f(\epsilon, \mu_L) = f(\epsilon, \mu_R)$
- A level that is at lower energy than both Fermi levels μ_L and μ_R will have $f(\epsilon, \mu_L) = f(\epsilon, \mu_R) = 1$ and will not contribute to the current.
- A level that is at higher energy than the μ_L and μ_R and has $f(\epsilon, \mu_L) = f(\epsilon, \mu_R) = 0$ will not contribute also to the current.
- It is only when the molecule's level lies between μ_L and μ_R (or within a few k_BT of μ_L and μ_R) that we have $f(\epsilon, \mu_L) \neq f(\epsilon, \mu_R)$, and a current flows.

The maximum current that can flow through one molecular level is equal to, $I_{max} = \frac{2e}{\hbar}\frac{\Gamma_L\Gamma_R}{\Gamma_L+\Gamma_R}$. The figure 2.6 shows a typical $I-V$ characteristics of one energy level calculated from equation 2.26 for different parameters. It can be seen that, at small bias voltages, the current is zero because both μ_L and μ_R are above the molecule energy level. When the voltage becomes sufficiently high, the μ_R drops below the energy level and the current increases to the maximum value I_{max}. It is worth noting that, the size of the gap in the current-voltage curve is equal to $4|E_f - \epsilon_0|$. Ef is the Fermi level of the system (left lead + molecule + right lead) at equilibrium, i.e. $V = 0$.

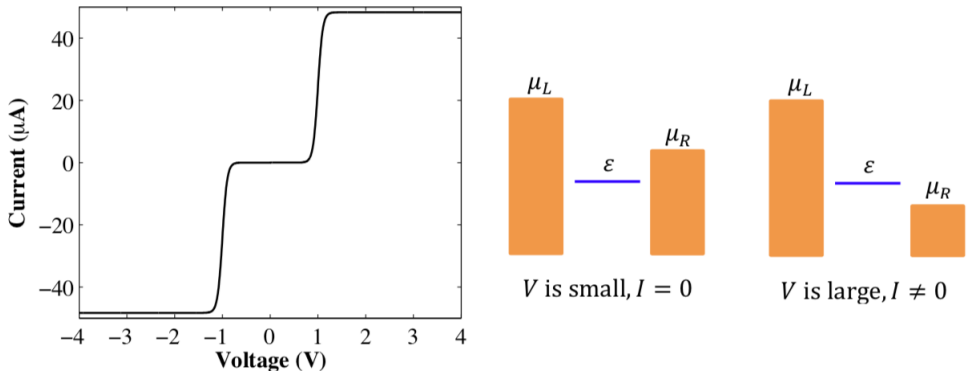


Figure 2.6: *(Right) The current-voltage (I-V) characteristics of a single energy level with $\mu_L = E_f - eV/2$, $\mu_R = Ev + eV/2$. (Left)The Voltage dependence position of the energy level relative to the μ_L and μ_R.*

When the current flows through the molecule of single energy level, electrons are added to or removed from the level continuously leading to what is called "charging effects". These effects have not been taken into account. The change in electron occupation of the energy level modifies the potential in the molecule. In order to make the model more realistic, we should include effects of charging.

Non-Equilibrium Green's Function and Landauer-Büttiker Formula

The simple model introduced in the previous section shed some light on basic factors that influence molecular conduction, but it does not describe two important aspects of a real device: the electronic structure of the contacts and the details of the scattering region Hamiltonian. These are replaced respectively by the Fermi distribution functions and by a single energy level. In reality, the molecular devices typically have multiple energy levels that broaden and overlap with each other and with the contacts levels. The non-equilibrium Green's function (NEGF) formalism does just that. The figure 2.7 shows a typical molecule in between two metal leads.

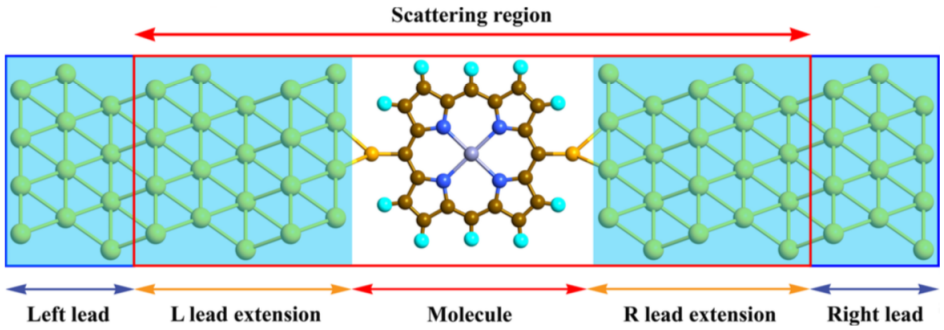


Figure 2.7: *molecule sandwiched between two leads.*

To investigate the electrical conduction through the molecule, the system is divided into three regions: the left lead, the right lead, and the scattering (central) region where the molecule of interest is located. The part of the scattering region where the atom positions follow the periodic arrangement of the leads is called the left and right lead extension. A sufficient fraction of the leads are included in the scattering region to screen out the perturbation of the scatterer, i.e. the molecule, in the outermost part of the scattering region. Since the two leads are semi-infinitely extended to the left and right sides (where the periodic boundary conditions are applied), the Hamiltonian of the whole system will be infinite in size. However, it can be computed by exploiting the periodic nature of the leads from smaller calculable components. To calculate the Hamiltonian of the system, for each lead a principal unit cell of calculation is chosen, which is at minimum the smallest number of atomic layers required so that no coupling exists between next nearest unit cells. As illustrated in Figure 2.8, each principal unit cell now has a finite Hamiltonian $\mathbf{H}_L$ and $\mathbf{H}_R$, and another matrix which describes the coupling between each principal unit cell and its next nearest unit $\mathbf{H}_{LC}$ and

$\mathbf{H}_{RC}$. The scattering region also has a finite Hamiltonian H_S and coupling matrices between itself and each lead $\mathbf{H}_{SLC}$ and $\mathbf{H}_{SRC}$.

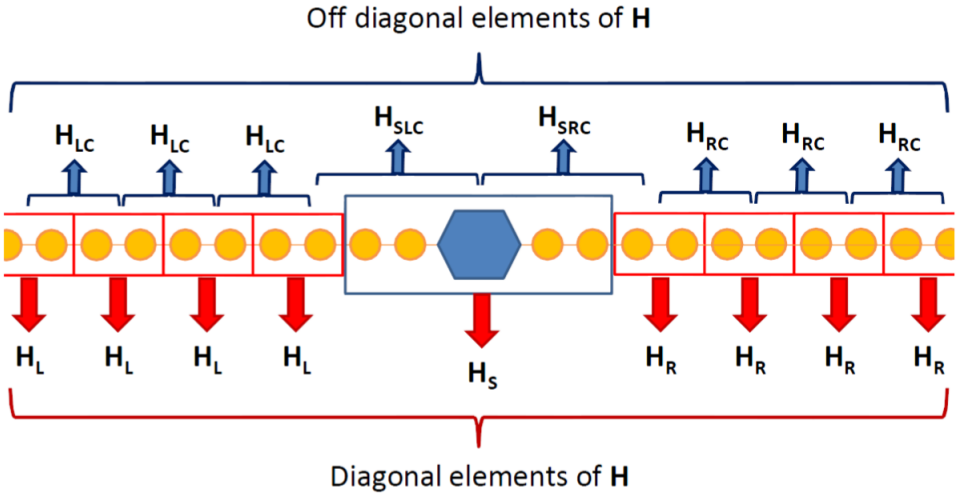


Figure 2.8: *Schematic diagram shows how the Hamiltonian of the one dimensional system is built up. In this example, each red rectangle, with two atoms, represents one principal unit cell of the lead, and the blue rectangle represents the scattering region which consists of left and right lead extensions and the molecule of interest.*

The Hamiltonian $\mathbf{H}$ describing the system can then be written as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_L & \mathbf{H}_{SLC} & 0 \\ \mathbf{H}_{SLC}^{+} & \mathbf{H}_S & \mathbf{H}_{SRC}^{+} \\ 0 & \mathbf{H}_{SRC} & \mathbf{H}_R \end{bmatrix} \tag{2.27}$$

The overlap matrix $\mathbf{S}$ can be obtained in the same manner; for a single energy level ϵ, the Green's function is defined as:

$$G(E) = (E - \epsilon + i\frac{\Gamma_L + \Gamma_R}{2})^{-1} \tag{2.28}$$

and in the case of the whole (infinite) system, the Green's function equation is given by

$$[(E + i\tau)\mathbf{S} - \mathbf{H}]\mathbf{G}(E) = \mathbf{I} \tag{2.29}$$

where τ is infinitesimal small positive number added to avoid singularity in calculation. I is the identity matrix with the same size of $\mathbf{S}$ and $\mathbf{H}$. What is required to calculate the current is the central block of the Green's function that corresponds to the scattering region (i.e. the molecule + leads

extensions). A reduced version of the last equation can be written in the following form:

$$[(E + i\tau)\mathbf{S}_S - \mathbf{H}_S - \Sigma_L - \Sigma_R]\mathbf{G}_S(E) = \mathbf{I} \tag{2.30}$$

$$\mathbf{G}_S(E) = [(E + i\tau)\mathbf{S}_S - \mathbf{H}_S - \Sigma_L - \Sigma_R]^{-1} \tag{2.31}$$

where Σ is a new matrix added such that the matrix $\mathbf{G}_S$ becomes equal to that results from equation 2.29; this new matrix is known as the leads self-energy matrix, and can be calculated (for left lead) as follows:

$$\Sigma_L = [(E + i\tau)\mathbf{S}^+_{LC} - \mathbf{H}^+_{LC}]\mathbf{G}_L[\mathbf{S}_{LC} - \mathbf{H}_{LC}] \tag{2.32}$$

where $\mathbf{G}_L$ represents the surface Green's function of the left lead, which is the bottom right block of the full semi-infinite Green's function representing the lead. The self-energy matrices, Σ_L and Σ_R, are generally complex energy dependence matrices containing all the details of the electronic structure of the leads and their coupling with the molecule. The broadening in molecule levels can be described using two matrices, Γ_L and Γ_R, which are defined as the "anti-hermitian" parts of the corresponding self-energy matrix, $\Gamma_{L(R)} = i[\Sigma_{L(R)} - \Sigma^+_{L(R)}]$. Knowing the Green's function of the scattering region allows us to calculate the current through a molecule by integration over energy spectrum:

$$I = \frac{2e}{h}\int_{-\infty}^{\infty} T(E)[f(E, \mu_L) - f(E, \mu_{L(R)})]dE \tag{2.33}$$

where $T(E)$ represents the electronic transmission through the scattering region at the particular energy E, $T(E) = Tr[\Gamma_L \mathbf{G}_S \Gamma_R \mathbf{G}^+_S]$. This is the voltage and temperature generalized Landauer formula [54] for an electric current flows through a molecular conductor between two electrodes. At low-temperature, the Fermi-Dirac distributions are reduced to step functions, and the integral in equation 2.33 is evaluated between μ_L and μ_R. Also, for low applied biases, the integral is confined to the vicinity of the Fermi level, so that $T(E)$ may be assumed to take the constant values $T(E_f)$. Under these conditions, the current can now be written as:

$$I = \frac{2E}{h} T(E_f)(\mu_L - \mu_R) \tag{2.34}$$

$$\mu_L - \mu_R = eV \tag{2.35}$$

$$\rho = \frac{1}{\pi} \int_{-\infty}^{\infty} dE \sum_{L(R)} f(E, \mu_{L(R)}) \mathbf{G}_S \Gamma_{L(R)} \mathbf{G}_S^{+} \tag{2.36}$$

$$G = \frac{I}{V} = \frac{2e^2}{h} T(E_f) = G_0 T(E_f) \tag{2.37}$$

where ρ is the electron density matrix and G is the conductance. In the case of a ballistic regime, i.e. when the mean free path of the electrons in the conductor is much less than the object's length, the transmission probability for all channels follows the equation 2.37 where the conductance will be a multiple of G_0; this implies that even a perfect conductor will exhibit ever a resistance. This resistance arises from the interface between the conductor and the contacts. The current is carried in the contacts by infinite number of channels but inside the conductor by only a few channels permit the transmission. Thus, the current is redistributed to a limited number of channels, causing a bottleneck [55].

2.3.2 Inelastic Condutance

While elastic transport properties give direct access to the electronic properties of the system (position and broadening of molecular levels for example), the study of the inelastic processes provides an additional tool to gain a further insight into the structure and dynamics of the molecular junction. When electrons tunnel through molecular junctions they can exchange (lose or absorb) energy with the vibrational modes of the molecule [51]. These processes can affect not only the conductance of the junction (hamper or enhance the current flow) but they can also modify its structure through for example the breaking or formation of bonds. The inelastic signal originates from the interaction of the electronic structure with the phonons. The inelastic effect occurs at electron energies above a corresponding phonon energy. An excitation of a phonon can only occur when the bias is equal to or above the phonon frequency, $eV < \hbar\omega$. The electron-phonon coupling can be described using the extended Hamiltonian equation:

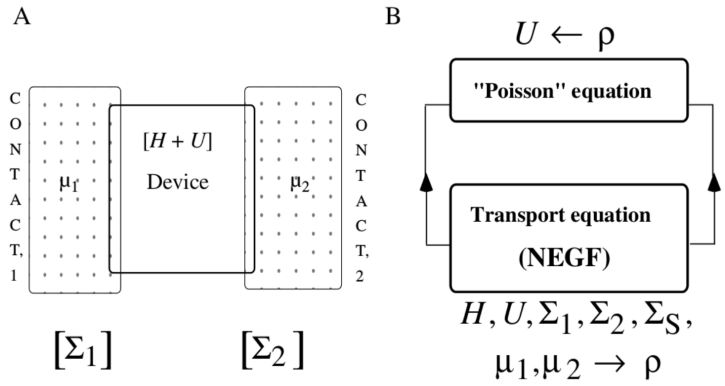


Figure 2.9: *A) The system consisting of the central device connected to two semi-infinite leads and B) the procedure scheme to analyze the system. When describing of the systems within the NEGF formalism, we can assign so-called self-energies to the contacts, which contain all the information about the coupling to the central device and how they change the energy of it. Together with the Hamiltonian of the central device - which we can obtained from DFT - we can then perform the self-consistent loop in figure B in order to calculate the electron density matrix ? of the system, which contains information about the electronic structure of the system. And with this, we can calculate all the quantities of interest, such as current, conductivity and transmission.*

$$\mathbf{H} = \mathbf{H}_e + \mathbf{H}_{ph} + \mathbf{H}_{e-ph} \tag{2.38}$$

$$\mathbf{H}_{ph} = \sum_{\lambda} \hbar\omega_\lambda \hat{b}^+_\lambda \hat{b}_\lambda \tag{2.39}$$

$$\mathbf{H}_{e-ph} = \sum_{\lambda} \sum_{i,j} \mathbf{M}^\lambda_{i,j} \hat{c}^+_\lambda \hat{c}_\lambda (\hat{b}^+_\lambda + \hat{b}_\lambda) \tag{2.40}$$

with $\mathbf{H}_e$ being the electronic Hamiltonian under the Born-Oppenheimer approximation [45]. $\hat{b}$, $\hat{c}$ are the phonon and electron annihilation operators ($^+$ creation). λ is the associated phonon index and i, j are the electron indices. $\mathbf{M}^\lambda$ describes the coupling element between the electronic structure and the phonon mode λ. The electron phonon coupling is a function of the electronic Hamiltonian derivative with respect to atomic coordinates

$$\mathbf{M}^\lambda = \sum_{Iv} \langle i| \frac{\partial \mathbf{H}_e}{\partial \mathbf{R}_{Iv}} |j\rangle v^\lambda_{Iv} \sqrt{\frac{\hbar}{2M_I\omega_\lambda}} \tag{2.41}$$

where v^λ_{Iv} is the normalised eigenvector for the phonon mode λ and frequency ω_λ. The following discussion will be based on the Lowest Order

Expansion Wide Band Approximation (LOE-WBA) and the LOE expansion [56]. The phonon modes and electronic Hamiltonian derivatives are calculated via the frozen phonon method [57]. The LOE-WBA approximation describes the inelastic current (without elastic contribution) for a single phonon λ as

$$I^{LOE-WBA} = y_\lambda I^{sym}_\lambda + k_\lambda I^{asym}_\lambda \tag{2.42}$$

$$y_\lambda = Tr[\mathbf{G}^+\Gamma_L\mathbf{G}(\mathbf{M}^\lambda\mathbf{A}_R\mathbf{M}^\lambda + \frac{i}{2}(\Gamma_R\mathbf{G}^+\mathbf{M}^\lambda\mathbf{A}_R\mathbf{M}^\lambda - H.c.))] \tag{2.43}$$

$$k_\lambda = Tr[\mathbf{G}^+\Gamma_L\mathbf{G}(\Gamma_R\mathbf{G}^+\mathbf{M}^\lambda(\mathbf{A}_R - \mathbf{A}_L)\mathbf{M}^\lambda + H.c.)] \tag{2.44}$$

where $\mathbf{A} = \sum_c \mathbf{A}_c$. The symmetric and anti-symmetric functions I, are respectively

$$I^{sym}_\lambda = \frac{e}{\pi\hbar}(2eV\langle n_\lambda\rangle + \frac{\hbar\omega_\lambda - eV}{e^{(\hbar\omega_\lambda - eV)/kT} - 1} - \frac{\hbar\omega_\lambda + eV}{e^{(\hbar\omega_\lambda + eV)/kT} - 1}) \tag{2.45}$$

$$I^{asym}_\lambda = \frac{e}{2\pi\hbar}\int_{-\infty}^{\infty} d\epsilon[n_F(\epsilon) - n_F(\epsilon - eV)]H_{\epsilon'}(n_F(\epsilon' + \hbar\omega_\lambda) - n_F(\epsilon' - \hbar\omega_\lambda))(\epsilon) \tag{2.46}$$

where $H_{\epsilon'}\epsilon'$ is the Hilbert transform. The above equations governing the LOE-WBA requires the Green function, spectral function and scattering matrices at the Fermi level. The wide band approximation is that the electronic structure is slowly varying in the limited energy range of the phonon spectrum. For systems with significantly varying density of states around the Fermi level an inclusion of the changing electronic structure needs to be taken into account. The LOE approximation improves the WBA by taking into account the change in electronic structure at the bias threshold of the phonon frequency [58]. The LOE estimates the conductance above threshold as a constant contribution (linear in current). The main differences can be outlined schematically in Figure 2.10.

The two electrodes, Left and Right, are shown with the associated applied bias eV. The DOS spectrum is marked in each of the two electrode regions by the thick black line. LOE-WBA and LOE are highlighted in different colours and the phonon excitation is shown as a wiggly line, $\hbar\omega$. An inelastic conductance contribution for phonon mode λ is constant for $\hbar\omega_\lambda \geq eV$ for both LOE-WBA and LOE. For LOE-WBA this contribution is for the spectral densities evaluated at $\epsilon = \epsilon_F$ (single line crossing both DOS's at same energy), while for LOE it is calculated using the spectral densities at

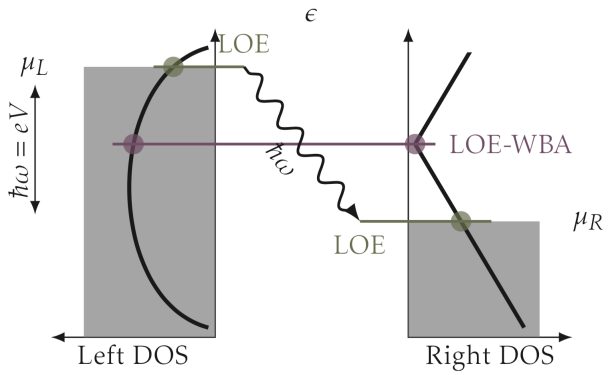


Figure 2.10: *Comparison of LOE-WBA and LOE contributions to the density of states from the two leads. The two methods have their own colour coding. The applied bias matches the phonon energy $\hbar\omega = eV$ which creates a constant contribution above $eV > \hbar\omega$ to the inelastic current in both approximations. The LOE-WBA only takes the Fermi level electronic structure into consideration. LOE takes into account the DOS from both the left (initial) state to the right (final) state.*

$\epsilon = \pm eV/2$ (different lines in each electrode). These two methods yields different results as the the DOS may change drastically, even for low phonon energies. However, the wide band approximation has proven sufficiently accurate in many cases. The LOE equations can be written without the elastic contribution as

$$I_\lambda^{LOE} = \frac{e}{\pi\hbar}\int_{-\infty}^{\infty} d\epsilon \mathbb{R}Tr[\mathbf{G}(\epsilon)\sum^{\lambda}(\epsilon)\mathbf{A}_L(\epsilon)\Gamma_R(\epsilon)](n_{F,L}(\epsilon) - n_{F,R}(\epsilon)) + \frac{e}{h}\sum_{\sigma=\pm}(coth\frac{\hbar\omega_\lambda}{2kT} - coth\frac{\hbar\omega_\lambda + \sigma eV}{2kT})\times \int_{-\infty}^{\infty} d\epsilon Tr[\mathbf{M}^\lambda\tilde{\mathbf{A}}_L(\epsilon)\mathbf{M}^\lambda\mathbf{A}_R(\epsilon_\sigma)](n_{F,L}(\epsilon) - n_{F,R}(\epsilon)) \quad (2.47)$$

where the threshold dependent energies $\epsilon_\sigma \approx \mu_R = \mu_L + \sigma\hbar\omega_\lambda$ are used instead of the constant Fermi level as in LOE-WBA. The phonon self energy, $\sum_\lambda$, can be found elsewhere [58]. The inelastic currents are seen as discontinuities in the current which can be understood via a constant DOS model. Here the elastic current $I = \int G$ is linear plus jumps at $eV = \hbar\omega_\lambda$. Hence, dI/dV yields a constant elastic conductance and a step for the inelastic conductance. A second differential d^2I/dV^2 only retains the inelastic contribution as a peak. The second derivative of the current is commonly known as the inelastic tunnelling spectroscopy signal (IETS) which may be

divided by the conductance, dI/dV , to decouple strong vs. weak signals via rescaling. In effect, IETS is the strength of the signal. Both the non-scaled and the scaled are seen in literature [56].

Inelastic Electron Tunneling Spectroscopy

In the field of molecular electronics, the widely employed experimental and theoretical techniques to study inelastic effects allow to uniquely characterize the compositional and structural details of system with unprecedented resolution [59]. Also, understanding the effects of electron-vibration coupling on charge transport through molecular junctions is a fundamental step toward the prediction and control of all those processes related to energy transport and conversion at the nanoscale (for example heating, cooling and thermoelectric effects). The interactions of electrons with molecular vibrations can have a strong impact on the current depending on the regime in which electron transport takes place. Typically, in the regime of weak electron-phonon coupling and strong coupling to the leads, a small correction to the elastic current is expected (in the order of few percent). In the complementary case (strong electron-phonon coupling and small coupling to the electrodes) a relatively strong change in the conductance is observable. A rule to separate the two regimes is to compare the strength of the electronic coupling of the "contact region" to the electrodes Γ to the phonon energy $\hbar\omega$ and to the electron-phonon coupling λ. In Figure 2.11 (a) are shown the different parameters defining the molecular junction.

If the "contact region" is strongly coupled to the electrodes ($\Gamma \gg \lambda$), the lifetime of the electronic states inside the molecule ($\tau_{ele} \propto 1/\Gamma$) is relatively small. This means that electrons have a short time to interact with molecular vibrations and, if λ is small (large time of interaction ($\tau_{int} \propto 1/\lambda$), this results in a small correction to the current. This is the case depicted in Figure 2.11 (b). In this regime inelastic electron tunneling spectroscopy (IETS) [60–62] provides a powerful tool for the identification of molecular adsorbates and junction geometry. The working principle is illustrated in Figure 2.11 (a)-(b). When an electron tunnels through the molecule, if it has enough energy and there are available density of states in the opposite electrode, it can release a quantum of energy $\hbar\omega$ to the vibrational mode. This excitation results in a change of the electrical conductance of the molecular junction; since this change is usually small, it could be appreciated looking at the second derivative of the current with respect to bias. In addition to the case shown in Figure 2.11 (b), where the inelastic process induces an increase in the differential conductance, for highly conducting junctions such as the case of metallic atomic wires, phonon excitations result in a decrease of conductance

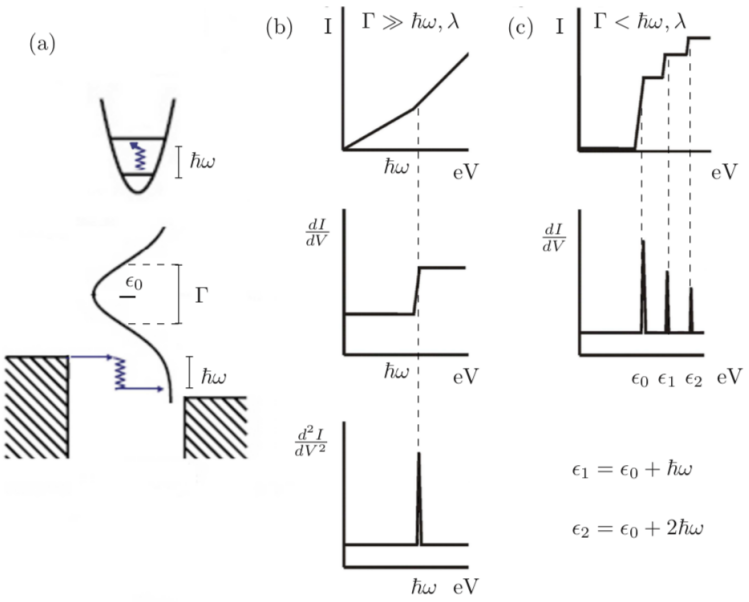


Figure 2.11: *(a) Illustration of the inelastic electron tunneling process through a single level model junction. In (b) is shown the current and its first and second derivative in the case of strong coupling to the electrodes and weak electron-phonon coupling. The emission/absorption of a phonon appears as a peak in the second derivative of current. In (c) is shown the current and the differential conductance in the vibronic regime. Multiple vibronic excitations can be observed as peaks in the first derivative of the current.*

[56, 63]. In the case of weak coupling to the leads and strong electron-phonon coupling the vibronic excitations can be observed directly in the I-V trace (Fig 2.11 (c)). Now, in fact, electronic states inside the "contact region" have a relatively long lifetime (small electronic broadening) and the time needed to interact with molecular vibrations is smaller (large λ). In the I-V curve the larger step is due to the electronic resonance entering the Fermi window and corresponds to a relatively large peak in the differential conductance while the smaller steps in the current are due to the vibronic excitations and result in smaller side peaks in the dI/dV curve.

Chapter 3

Atomistic Modelling

In this chapter, we describe the choices and motivations of the systems adopted and of the protocols used for the simulations. For the classical simulations, was chosen the biological nanopore of the transmembrane protein α-Hemolysin and for target of the homo-peptidic sequences; for quantum simulations, was chosen a solid-state nanopore of graphene and for target both of the homo-peptidic and of the hetero-peptidic sequences.

3.1 Classical Molecular Dynamics

Molecular Dynamics with NAMD

NAMD (NAnoscale Molecular Dynamics) [64] is a MD code suitable for simulating nanopore translocation. The code is designed for high-performance large-scale simulations on parallel processors, which makes it ideal to simulate the time-evolution of a large atomistically detailed model such as that of nanopore translocation. NAMD use the CHARMM force-fields [65]. In an atomistically detailed MD simulation, every atom is described by a model force-field which accounts for the interactions between the atom and the system. NAMD uses contributions to the potential energy function from the following interactions:

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{Coulomb} \tag{3.1}$$

U_{bond}, U_{angle} and $U_{dihedral}$ describe the stretching, bending, and torsional bonded interaction respectively. U_{vdW} describes the non-bonded Van der Waal's forces, which is approximated by a Lennard-Jones potential. $U_{Coulomb}$ describes the non-bonded electrostatic interactions. NAMD allows the use of periodic boundary conditions to avoid surface effects at the boundary of

simulated systems. The particles in the system are subject to the potential of all the atoms in the cell as well as the potential of the atoms in the repeating cells. The influence of potential from repeated cells in the system is only of concern for the van der Waals forces U_{vdW} and electrostatic integrations $U_{Coulomb}$, and are computationally unfeasible to calculate the exact long-range interactions. U_{vdW} is approximated by spatially truncating the interactions at a user-defined cutoff distance. For $U_{Coulomb}$, the particle-mesh Ewald (PME) system is applied. PME interpolates the charges of the system to the nodes of a three dimensional grid, so that when a charge is considered by the system, the electrostatic interactions are calculated with respect to the node points which are given weighting functions according to the distance to the charge. This fast numerical method allows the long-range electrostatic interactions to be calculated.

3.1.1 System Chosen

In this section, we use of the biological-nanopore formed by α-Hemolysine and four homo-polypeptides for target, we analyze the behavior of the peptide in the pore in presence of electric field, this provides of the data to know the phenomena of interaction between the peptide and the pore; this through analysis of current blockage induced by peptide within the pore. All-atom Molecular Dynamics (MD) simulations were performed using the NAMD software [64], the CHARMM36 force field [65] was employed to model lipid, protein, and TIP3P water molecules [66], NBFIX corrections were applied for ions [67].

Device and Target

We studied the ionic current for four different homopeptide chains clogging the αHL nanopore via all-atom molecular dynamics simulations. The four amino acids composing the homopeptides are alanine (Ala), phenylalanine (Phe), tryptophan (Trp) and glutamine (Gln) each composed of 35 residues, for each of them we prepared five independent replicas; the system set-up is sketched in figure 3.1.

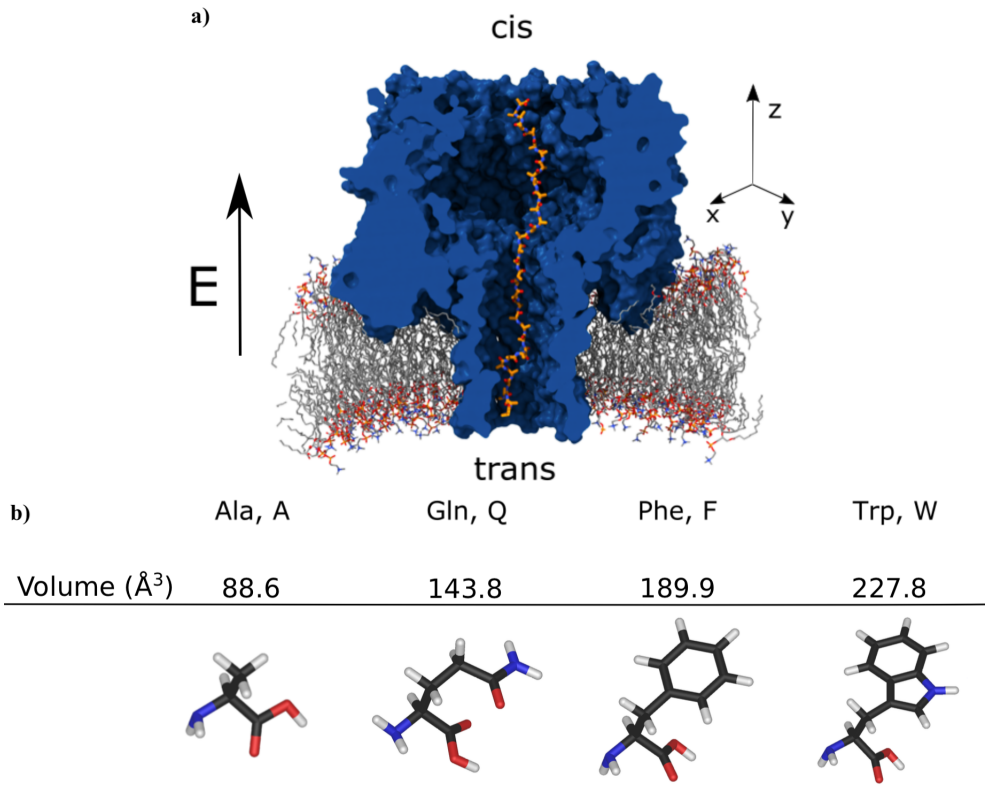


Figure 3.1: *a) The system is constituted by an α-Hemolysin (αHL, blue) nanopore embedded into a lipid membrane (red). A 35-residues homopeptide (orange chain) is imported into the nanopore with the central residue close to the pore constriction. The simulation box is filled up by 2M KCl electrolyte solution, that, for sake of clarity, is not shown. A constant and homogeneous external electrical field E = (0, 0, E_z) parallel to the pore axis is applied. b) Molecular structure and Van der Waals volume of the four amino acids.*

The membrane and α-Hemolisine (αHL) system has been assembled according to protocol [68]. The system was built starting from the αHL crystal structure PDB ID: 7AHL33 downloaded from the OPM database [69]. The POPC lipid membrane, the water molecules, and the ions for neutralizing the system were added using VMD [70]. Then, the system is minimized and a 60 ps NVT simulation (time-step 0.2 fs) was run with external forces applied to the water molecules to avoid their penetration into the membrane and the pore. Lipid heads have been constrained to their initial position by means of harmonic springs (spring constant k = 1 kcal/(molÅ^2)) acting on the phosphorus. A second equilibration run (1 ns NPT flexible cell, time-step 1 fs) was performed to compact the membrane. During this run, the lipid

heads were unconstrained. The third, and last, equilibration step consists of a NPT constant area simulation (2 ns, time step 2 fs) where all the atoms are unconstrained and no external forces act on the water molecules. The resulting periodic box, after the equilibration, has the following size: L_x = 127.5 Å, L_y = 127.1 Åand L_z = 180.0 Å, and the total number of atoms is $\sim$ 290000. Initial configurations of peptides are generated by using the PEPFOLD server [71] and then separately equilibrated in a triperiodic water box. Then, the two systems were merged, ions (2M KCl) were added using VMD, and a short NPT equilibration is performed (2 ns, constant area NPT) until Lz reaches a stationary value. The resulting box has dimensions L_x = 127.5 Å, L_y = 127.1 Å(i.e. the same as the original equilibrated αHL-membrane box) while $L_z \simeq$ 186.2 Å(slightly different values are getting for each homopeptide) and the overall number of atoms is $\sim$ 310000.

3.1.2 Simulation Set-Up

After equilibration for each replica, the homopeptide is imported into the pore using Steered Molecular Dynamics [43] to bring the peptides at the pores lumen entrance (trans side) and, then, into the pore. In particular, the peptide N-terminus was placed at $\sim$ 15 Åfrom the αHLs trans entrance and then pulled inside the nanopore using a constant velocity Steered Molecular Dynamics (SMD) simulation at pulling speed v_{SMD} = 0.025 Å/ps. We then select the frame for which the C_α of the central residue of the homopeptide is closer to the pore constriction defined as the average position of the seven copies of amino acid (AA) of the αHL heptamer.

This configuration was used for non-equilibrium runs where uniform and constant external electric field E = (0, 0, E_z) was applied perpendicularly to the lipid bilayer, corresponding to a trans membrane voltage ΔV = 1V; this protocol was shown to be equivalent to the application of a constant voltage $\Delta V = E_z L_z$ [72]. Each simulation was run for 240 ns and snapshots are saved every Δt = 40 ps. The average current in the interval [t, t + Δt] is estimated as (formula 2.4):

$$I(t + \Delta t/2) = \frac{1}{\Delta t L_z} \sum_{i=1}^{N} q_i (z_i(t + \Delta t) - z_i(t)) \tag{3.2}$$

where q_i and z_i are charge and the z-coordinate of the i-th atom, respectively. The K^+ and Cl^- currents were computed by restricting the sum over the atoms of corresponding type [68]. The mean current is obtained via a time average of $I(t)$ after discarding a transient of 64 ns. For each one of the four homopeptides, we have sampled 5 independent series of data, for a total

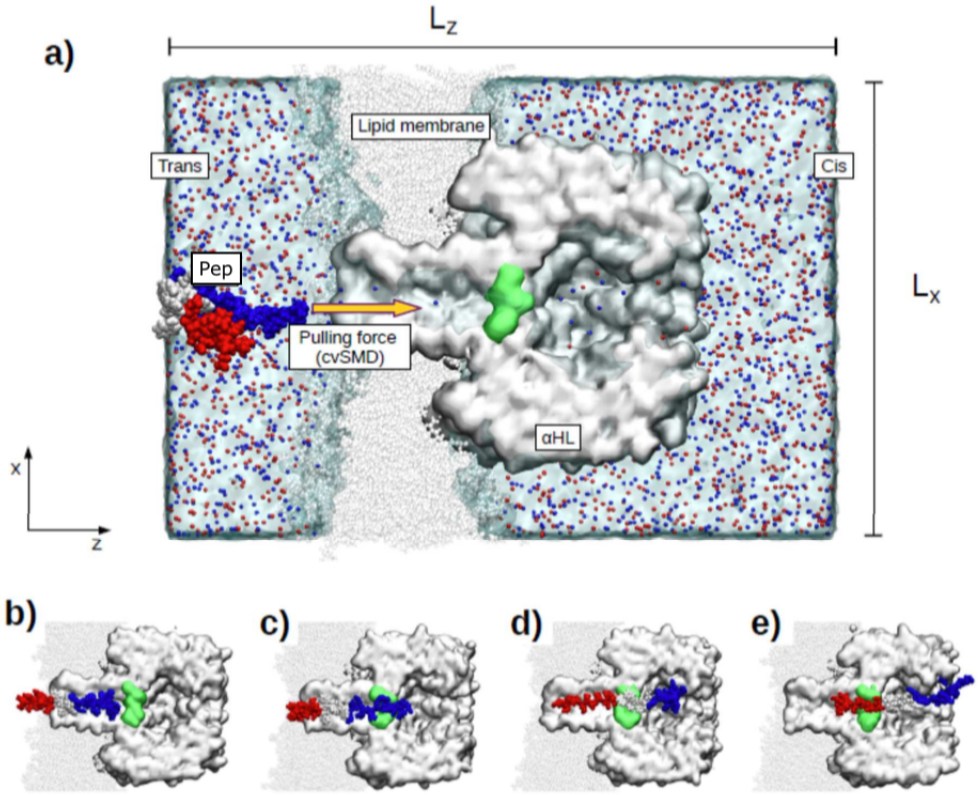


Figure 3.2: *a) The system is constituted by the α-HL nanopore embedded in a lipid bilayer, the peptide, water and ions (2M KCl). The α-HL's constriction (Glu 111, Met 113 and Lys 147) is highlighted in green, while, the peptide residues are colored according to their "zone", i.e. the last 11-residues are in red, the first 11-residues are in blue and the middle 13-residues are in white. The snapshot represents a configuration for the peptide after the complete equilibration procedure. The peptide is already imported at the nanopore's barrel (trans) entrance. The translocation is induced by a constant velocity Steered Molecular (cvSMD) dynamics, i.e. a harmonic spring links the alpha-carbon of the C-terminal residue to a ghost atom that moves in the z-direction with constant velocity. Panels b-e report some intermediate steps along the translocation pathway. The configuration d was used for non-equilibrium runs where uniform and constant external electric field E = (0, 0, E_z) was applied.*

of 20 ionic current tracks. Each series is composed by N = 4400 uncorrelated current data points corresponding to the average current value over 40 ps intervals. To evaluate if the average ionic current I measured for one replica is significantly different from another one, a series of Student's t-test (null hypothesis: the average current is the same) has been conducted. Figure 3.3

reports the p-value matrix of t-test between each couple of samples. Red boxes indicates $p < 0.05$, while light blue ones refer to $p > 0.05$. All tests are computed using R2.

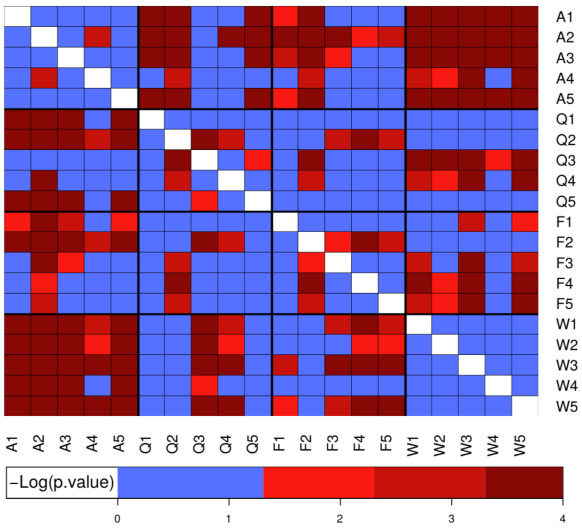


Figure 3.3: *Each row (and column) refers to a single replica, where, following the same nomenclature of figure 1c of the main text, the letter corresponds to the amino acid forming the homopeptide and the number refers to the replica. Red squares represent significant differences, adjusted $p < 0.05$, while light blue ones refer to non significant differences (adjusted $p > 0.05$).*

3.2 Quantum Molecular Dynamics

DFT in TranSIESTA

The *ab-initio* results shown in this thesis are obtained with the DFT code SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) [73] which has been extended for the treatment of quantum transport (TranSIESTA) [74] with the NEGF. In DFT calculations solids and nanostructures are often represented as a periodic repetition (in the three directions of space) of a relatively small unit called supercell. The solution of the Schroedinger equation assuming these periodic boundary conditions satisfies Bloch's theorem [75] since it can be written as the product of a plane wave and a periodic function with the same periodicity of the crystal:

$$\Psi_{j,k}(r) = e^{ikr} u_{j,k}(r) \tag{3.3}$$

where j is a discrete band index and k the reciprocal lattice vector. In SIESTA, the sampling over the reciprocal space is based on the so-called Monkhorst-Pack scheme [76] which allows to reduce the number of k-points needed. Many quantities, such as the density of states (DOS), thus require to be integrated over the first Brillouin zone. The relevant aspect of this code for the treatment of quantum transport is that the basis functions are localized in space, which is a necessary requirement for the implementation of the NEGF scheme. The eigenfunctions are in fact expressed as a linear combination of atomic orbitals (LCAO) centered at the atoms positions. Such orbitals are strictly confined in space by a cut-off radius which is indirectly defined in terms of a confinement energy. The basis functions are expressed as a product of a radial function and a spherical harmonics:

$$\psi_{I,l,m,n}(r) = R_{I,l,n}(|r_I|) Y_{l,m}(\hat{r}_I) \tag{3.4}$$

where $R_{I,l,n}(r_I)$ and $Y_{l,m}(r_I)$ are the radial and the angular part respectively; I defines the position of the atom (with $r_I = r - R_I$), l and m are the shell and orbital angular momentum while the subindex n indicates that, for each l, more radial functions can be used. In a minimal basis definition (single-ζ) there is just one radial function for each shell; a more complete basis can include more radial functions for each shell (multiple-ζ) and additional shells with different spatial symmetry (polarization). The chemical inert core electrons are described by an effective potential called pseudopotential which is parametrized according to the Troullier-Martins method [77]. This reduces notably the computational cost of the calculation since only valence electrons are explicitly considered. Overlap matrix and most elements

of the Hamiltonian are calculated in the energy domain while the remaining terms and the charge density are calculated over a real space grid. The fineness of the grid is another input parameter called "grid-cutoff".

3.2.1 System Chosen

DFT-NEGF, is nowadays the most popular approach for steady-state transport in nanostructure and has been successfully applied in many cases. DFT-NEGF provides a correct qualitative insight on the transport behavior, and often gives current values close to the ones obtained with formally correct methods [78, 79]. Using the NEGF formalism together with the DFT method, one can now study systems like a graphene nanoribbon (GNR) with an embedded nanopore [80]. An illustration of the system

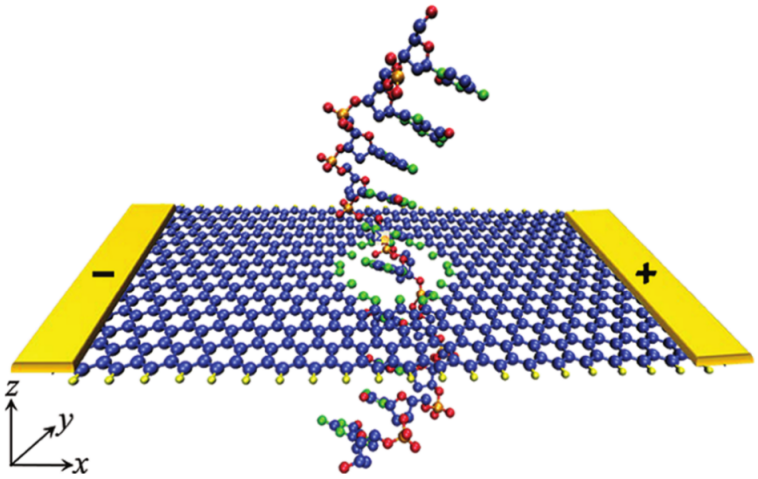


Figure 3.4: *Graphene nanoribbon with an embedded nanopore, a DNA molecule in the middle of the pore, and the two electrodes on the sides.*

This set up of graphene nanoribbon is with zigzag edges (ZGNR) with an embedded nanopore in the middle of the ZGNR. The edge atoms of the ZGNR are passivated by hydrogen. The question now is, how it is possible to perform sequencing with such a pore. The key to that is the so-called "modulation of edge currents". It is known that in ZGNRs, the electrons at the edge of the ribbon are the main factor when it comes to conductance. When something enters the pore, its presence will affect the electronic states of the edge atoms, thus modulating the conductance and thereby the current. Similar simulations can be performed with a graphene nanogap as illustrated in figure 3.5.

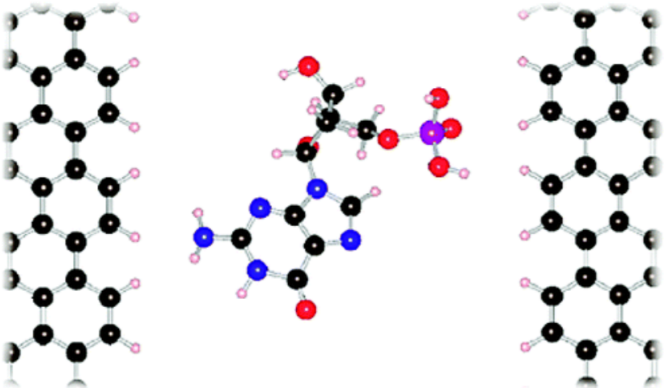


Figure 3.5: *Illustration of the graphene nanogap, as studied by Jariyanee Prasongkit et al. [81]. The two ZGNRs with a gap width of 1.47nm contain a single nucleobase in their center while the tunnelling current is being measured.*

The system consists of two ZGNRs, the edge carbon atoms are passivated by hydrogen. Unlike for the system with the graphene nanopore, where we measured the modulation of the edge currents, we now measure the tunnelling current through the nanogap; where the calculation of the Transmission function, tells us about which energy "channels" the electrons can use in order to move through the material, thus conducting a current.

Device and Target

The atomistic model we adopted is basically made of two semi-infinite graphene nano-ribbons with zig-zag edges (ZGNR) and a peptide that translocates across the nano-gap. The chosen ZGNR is characterized by a gap in the transmission function around the Fermi energy with zero dark current.[82] The idea is to exploit the ZGNR conductance properties to detect current signals from the peptide bonds and the amino acid side chains. Post-processing of these current signals would lead to the recognition of amino-acids (AAs) or, at least, of the AA types (neutral, polar, charged, small, large and so on). The chains analyzed are of two types: homopeptides and heteropeptides each composed of 4-residues; the homopeptide is Glycine chain, while the heteropeptides are formed by Glycine alternating with Alanine, Aspargina and Aspartate.

DFT and quantum transport calculations have been carried out in a dry environment since the conductance of ZGNR nanoribbons does not vary significantly when the dynamical environment of water is removed from the system [84] due to the hydrophobic character of graphene-based structures. Moreover, because of the narrow size of the ZGNR nanogap, water molecules

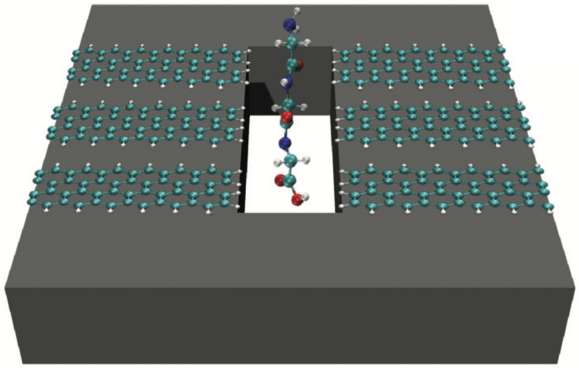


Figure 3.6: *The proposed device is constituted by an array of GNR nano-gaps deposited onto a substrate with a nanopore. Tunneling current signals could be collected between the electrodes depending on the size and chemistry of the AAs. The central nano-gap should be devoted to collect the PB related signals [83].*

cannot be allocated between the leads together with the PBs but are expected to cross the nanodevice near the lateral ZGNR nano-gaps where the AA side chains move across the device. To say this, we have studied the transport properties of the nanostructures considered taking into account the possible influence of water molecules. First, we have considered the case of the ZGNR with a close layer of water molecules. This configuration has been fully relaxed, in order to get the ground state configuration that, in principles, should be the ones with water molecules better adapted close to the GNR. In Figure 3.7(a) we report the transmission function in the relevant energy range of the hydrated ZGNR as compared to the dry one: the two curves are practically indistinguishable. As a second issue, we have optimized the configuration of one single water molecule in the middle of the nano-gap. we have found that in the ground state configuration, where also the current signal is maximum, the water molecule is just above the ZGNR plane. The transmission function and the current measured in this case have been compared to one of the global minima of the PB current signal measured for a Gly homo-peptide, as reported in Figure 3.7(b). We see that the signal measured in this case is well below the minimum transmission function obtained with the Gly side chain in the nano-gap. The measured current value is $I_{SC} \approx 0.15$ nA that is about one order of magnitude lower than the maximum signal measured in the Gly homo-peptide and well below the absolute minimum signal with the side chain in the gap $I_{SC} \approx 0.4$ nA.

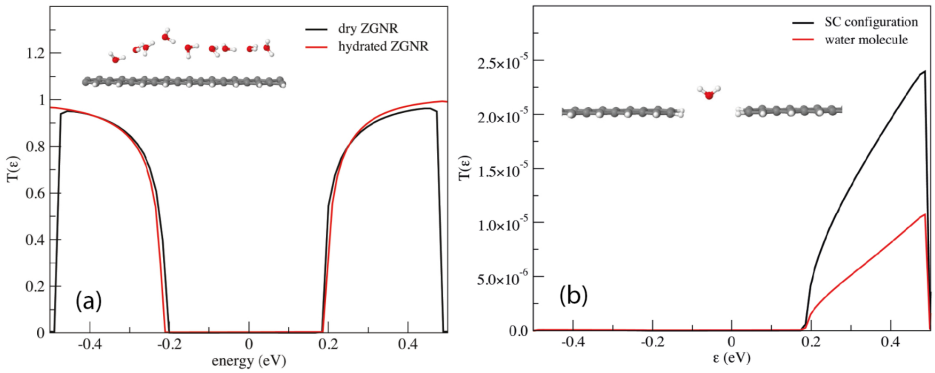


Figure 3.7: *Transmission function of the hydrated ZGNR as compared to the dry one (a) and of the optimized configuration of a ZGNR nano-gap with a water molecule in between as compared to the one corresponding to the minimum current value measured in the Gly homo-peptide (b). The atomistic models of these optimized configurations are shown in the insets.*

3.2.2 Simulation Set-Up

Following the scheme of Figure 3.8, in our calculations each lead has been modeled as a semi-infinite symmetric graphene nanoribbon with zig-zag edges (ZGNR), 5.0 Åwide and lying in the z direction. The nano-gap has a hydrogenated armchair pattern at the border and the distance between the two semi-infinite ZGNRs is 5.0 Å(distance between H atoms).

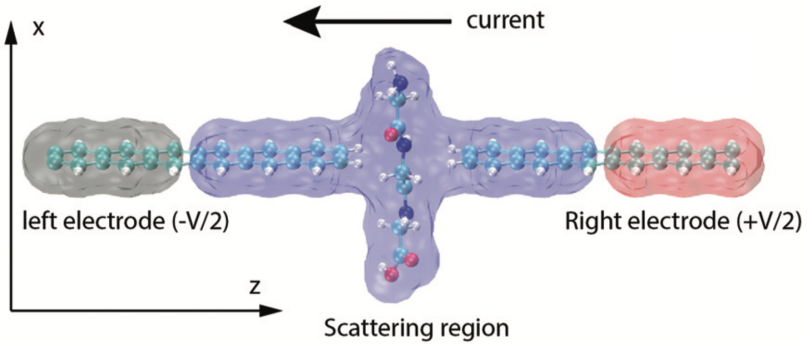


Figure 3.8: *The atomistic model adopted in the present work to calculate the current using the NEGF linear approach; C, O, and N atoms are respectively in cyan, red, and blue. The gray, red and blue haloes represent respectively the left and right electrodes and the central scattering region.*

The configurations of the amino acids crossing the gap have been identified using NVT classical steered molecular dynamics (SMD) simulations [43]

at T = 300 K in a water environment. Then, conformations along the translocation pathway have been selected, water molecules have been removed and the ionic degrees of freedom of the system have been relaxed at T = 0 K, in the context of density functional theory (DFT) in order to approach a local energy minimum. The total energy DFT calculation scheme includes a generalized gradient approximation based on the Perdew-Burke-Ernzerhof formula [48] (PBE) for the electron exchange and correlation potential V_{XC}[n(r)] and norm-conserving pseudopotentials built with the Troullier-Martins scheme [77] in the framework of a plane-wave basis set expansion. DFT calculations have been performed using the Quantum ESPRESSO package [85], with a plane-wave energy cutoff of 70 Ry for the wave functions and an artificial smearing of 0.005 Ry, while structural optimization has been achieved in a box of 14.7×14.7×36.16 Å^3 using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [86] together with the Hellmann-Feynman forces acting on the ions; an empirical dispersion force field [87] has been included to handle the long-range interactions. The chosen box size is large enough to prevent any spurious interaction arising from the periodic boundary conditions (PBC); thus calculations have been performed using the Γ point for the Brillouin zone sampling, and the ionic minimization has been carried on until the convergence threshold of 0.001 a.u. for the total force was achieved. Quantum transport calculations have been performed using the non-equilibrium Green's function (NEGF) method as implemented in the TRANSIESTA code [74]. Again TM norm-conserving pseudopotentials have been employed, together with a DZP localized basis set. A 1×1×50 Monkhorst-Pack [76] k-grid was used for electrodes. The electronic temperature was set to 0 K. The transport properties of the ZGNR junction have been evaluated by means of the Hamiltonian decomposition (formula 2.27):

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_L + \Sigma_L & \mathbf{V}_L & 0 \\ \mathbf{V}_L^+ & \mathbf{H}_S & \mathbf{V}_R^+ \\ 0 & \mathbf{V}_R & \mathbf{H}_R + \Sigma_R \end{bmatrix} \tag{3.5}$$

where $H_{L(R)}$ and H_S are, respectively, the decoupled Hamiltonians of the left (right) electrode (composed of ZGNR) and the scattering region (composed of ZGNR + peptide), $V_{L(R)}$ represents the interaction of the left (right) electrode with the scattering region and $\Sigma_{L(R)}$ is the self-energy and describes the coupling of the left (right) lead with the scattering region. For an external bias voltage V applied along the z direction, the current $I(V)$ is given by the Landauer-B*ü*ttiker (formula 2.33) [54]. The current signals have been measured on each peptide for an applied voltage ΔV = 1 V between the leads. The present bias has been employed both in the theoretical and experimental literature concerning tunneling currents or molecular junctions

[88–90] but, of course, it should be tested for a breakdown in real systems, especially for GNR electrodes [91]. In any case, we have not observed any density or wave function anomalous behavior up to a bias of 2 V. In the end, for a deeper insight on the local transport phenomena occurring at the atomistic level, we have looked at the local currents flowing across the atomic sites [92]. Given a suitable complete localized basis set $|\phi_{n,\gamma}\rangle$ in a subspace of the full device region where $|\phi_{n,\gamma}\rangle$ is an atomic orbital of type γ, centered at the atomic site n, the local current flowing between atoms n and n' is [93]:

$$I = e \int de[f_L(\epsilon) - f_R(\epsilon)]Tr[G(\epsilon)\Gamma_L(\epsilon)G^+(\epsilon)J_{n,n'}] \tag{3.6}$$

$$J_{n,n'} = \frac{1}{i\hbar}(P_n H P_{n'} - P_{n'} H P_n) \tag{3.7}$$

where P_n is the projection operator on the n site. The atomic currents at the left lead (electron injection) have been calculated in the energy range [-0.5 eV $\leq \epsilon \leq$ 0.5 eV] with the INELASTICA package [94].

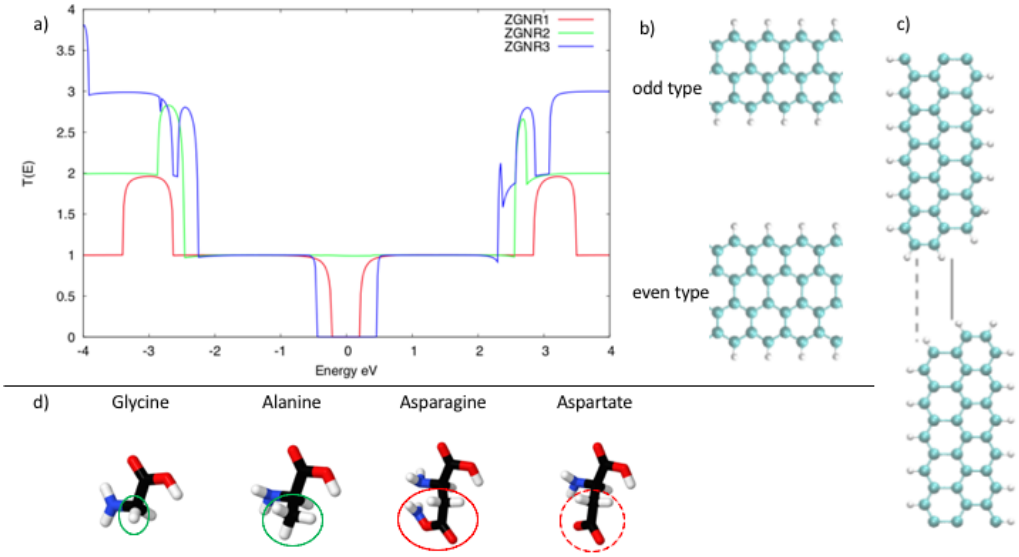


Figure 3.9: *a-b) Trasmission functions of zig-zag graphene nano-ribbons of even type (one and three rings: ZGNR1 and ZGNR3) and of odd type (ZGNR2). The transport properties, the chosen ZGNRs are of two types: ZGNR1 and ZGRN3 belong to the so called "even-type" GNRs with transport properties characterized by a gap in the transmission function around the Fermi energy and zero dark-current, while ZGNR2 is of the "odd-type" with a continuous positive transmission function around the Fermi energy and no gap [82]. This circumstance seems to indicate that odd-type GNRs should be preferred but when a gap is opened in the GNR the nano-gap size is characterized by two distances (example c) 5 Å(black line) 7.5 Å(dot line)) so to avoid problems with symmetry we initially chose a 1-ring GNR [95]. d) The peptides selected in the rings side-chain, green non-polar (Gly, Ala), red polar (Asn), red dot charge (Asp)*

Chapter 4

Results and Discussion

In this chapter we expose the results obtained using the theoretical methods described in chapter 2 used with the set-up described in chapter 3, the results are divided by the classical molecular dynamics where we calculated the ionic current 4.1 and for the quantum dynamics where we have elastic 4.2 and inelastic 4.3 signal.

4.1 Current Blockage

The current blockage is defined as $\Delta I/I_0 = (I_0 - I)/I_0$, with I the average current measured with the homopeptide inside the pore and I_0 the empty pore value. Figure 4.1b shows the average current blockage $\Delta I/I_0$ for each homopeptide, which are obtained pooling the data from the five independent replicas, while Figure 4.1c reports $\Delta I/I_0$ for each single replica. As expected, $\Delta I/I_0$ roughly reflects the steric hindrance of each amino acid. Indeed, the lower blockage corresponds to Ala (VdW volume [96], V_A= 88.6 Å^3) and the largest to Trp (V_W = 227.8 Å^3) while Gln (V_Q = 143.9 Å^3) and Phe (V_F = 189.9 Å^3) blockages are in between the Ala and Trp values. Interestingly, significant differences among replicas of the same homopeptide are found for Ala, Gln, and Phe, while, Trp replicas do not show any significant variability (see Section 3.1.2 and figure 3.3). This occurrence can be explained in term of the capability of smaller amino acids to explore a larger number of conformations inside the pore.

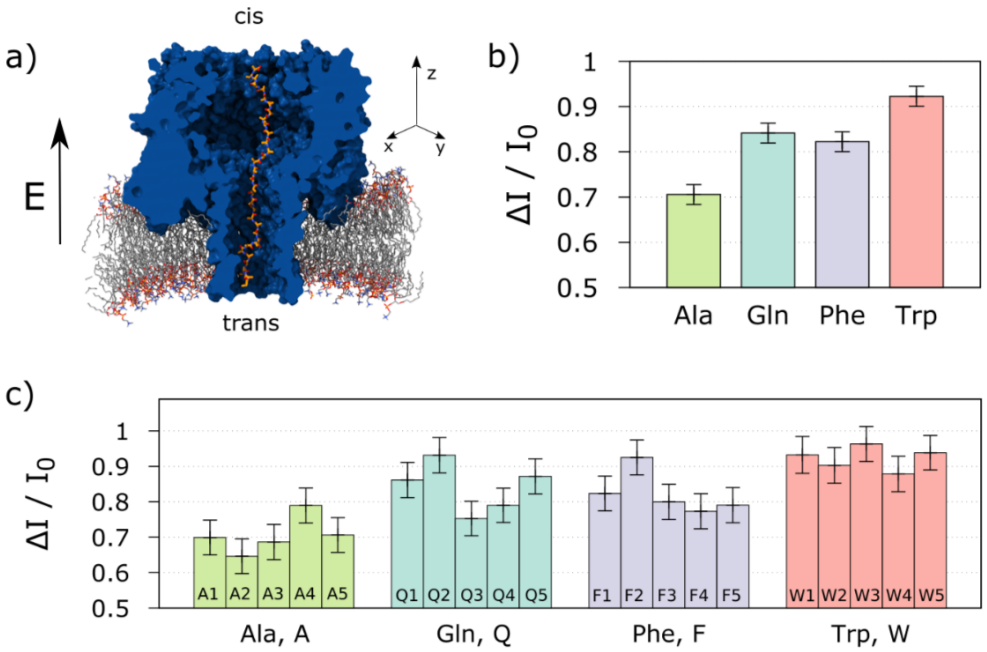


Figure 4.1: *a) The system is constituted by an α-Hemolysin (αHL, blue) nanopore embedded into a lipid membrane (red). A 35-residues homopeptide (orange chain) is imported into the nanopore with the central residue close to the pore constriction. The simulation box is filled up by 2M KCl electrolyte solution, that, for sake of clarity, is not shown. A constant and homogeneous external electrical field E = (0, 0, E_z) parallel to the pore axis is applied. b) Average current blockage $\Delta I/I_0 = (I_0 - I)/I_0$, with I the average current measured with the homopeptide inside the pore and I_0 the empty pore value, for four different homopeptides, Ala, Phe, Gln, Trp. The data are obtained pooling the current I from five independent replicas for each homopeptide. c) Current blockages for the single replicas, significant differences among replicas for the same homopeptide are found for Ala, Phe and Gln.*

The above presented results indicate that the size of the side chain is correlated to the current blockage; the larger the side chain, the deeper the current drop. To better investigate the role of peptide conformation on the current drop, we formulated the following simple theoretical model. In a quasi-1D continuum description, the pore resistance is expressed as

$$R = \int_0^L \frac{\rho(z)}{A(z)} dz \tag{4.1}$$

where the z-axis coincides with the pore axis, the pore goes from $z = 0$ to $z = L$, $\rho(z)$ is the electrolyte resistivity and $A(z)$ is the area of the pore

section available to the electrolyte. Access resistances are neglected. To estimate $A(z)$ from our non-equilibrium runs, we divided the system in cubic cells of size $\Delta x = \Delta y = \Delta z = 1$Å, and, for each frame, we used the VMD Volmap [70] to compute the occupancy map of the electrolyte, $m_{x,y,z}$, where x, y, z indicate the cell, $m_{x,y,z} = 1$ if the cells is within a Van der Waals radius of at least one water or ion atoms and $m_{x,y,z} = 0$ elsewhere. Then, we averaged $m_{x,y,z}$ over all frames and normalized it with the bulk value. The resulting averaged and normalized occupancy map is indicated with $M_{x,y,z}$. As already discussed in Aksimentiev and Schulten [68], "electrolyte pockets" are present close to constriction, see figure 4.2.

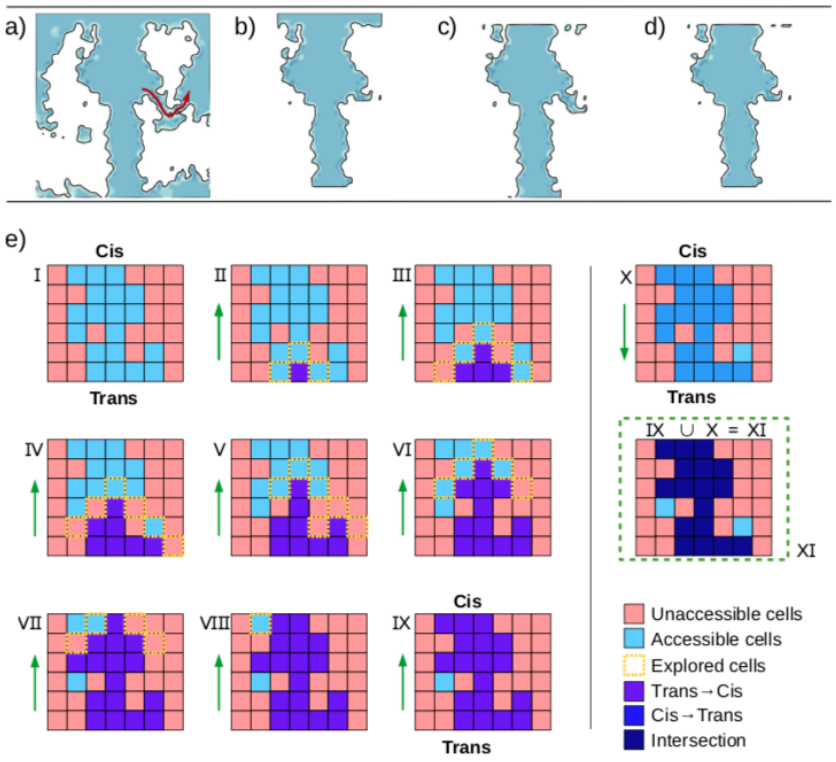


Figure 4.2: *a) Averaged and normalized map $M_{x,y,z}$. Blue areas cor- respond to region that are fully accessible by the electrolyte $M_{x,y,z} = 1$ while white ones correspond to $M_{x,y,z} = 0$. Water reentrant pockets, already observed in4 are apparent close to the constriction (red arrow in panel a). Although these pockets are connected to the electrolyte region inside the pore, they do not contribute to the electrolyte motion between the two sides of the membrane. To filter our these dead-end pockets, we set up sketched in panel (e), that allowed us to define a trans → cis and an cis → trans maps, panels (b) and (c). The intersection between these two maps leads to the final map $\tilde{M}_{x,y,z}$, panel (d).*

The pockets do not contribute to the ion current. To filter out these pockets, we defined a *trans* $\rightarrow$ *cis* available channel as the pore region accessible to the electrolyte when moving from the barrel entrance towards the vestibule. This procedure excludes reentrant pockets directed towards the trans side, see figure 4.2. The same procedure is applied to get a *cis* $\rightarrow$ *trans* accessible pore, and the final occu pancy map $\tilde{M}_{x,y,z}$ is obtained as the intersection of the *trans* $\rightarrow$ *cis* and *cis* $\rightarrow$ *trans* accessible pores. Figure 4.3c reports slices of $\tilde{M}_{x,y,z}$ for the four homopeptides. The regions available for the electrolyte transport between the two sides of the membrane are indicated in blue.

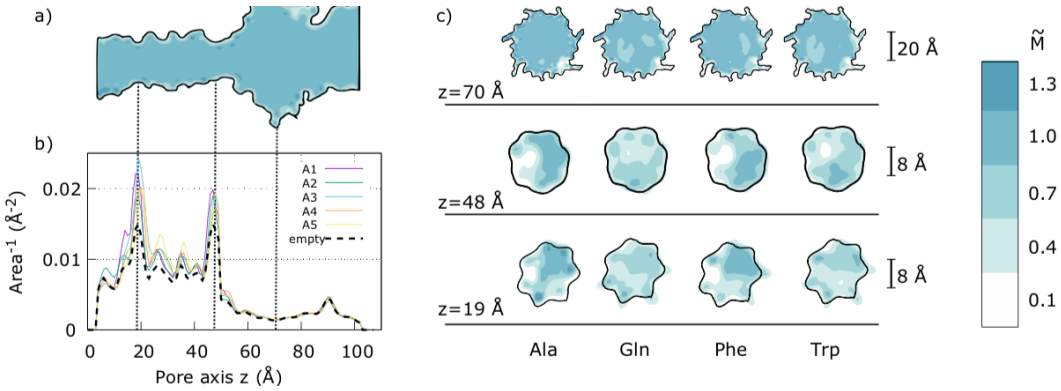


Figure 4.3: *a) The panel report a cut of the 3D averaged occupancy map $\tilde{M}_{x,y,z}$ for the empty pore. Blue areas corresponds to region that are fully accessible by the electrolyte $M_{x,y,z} = 1$ while red ones do not contribute to the volume useful for the ions transport between the two side of the membrane, $\tilde{M}_{x,y,z} = 0$. b) Inverse of the accessible area, A_z, along the pore. The empty pore profile is plotted as dashed black line. The two peaks at $z \simeq 50$ Åand $\simeq 20$ Åcorrespond to the central αHL constriction and to the constriction close to the barrel entrance, respectively. The five solid lines refer to the five Ala replicas. c) Slices of $\tilde{M}_{x,y,z}$ normal to the pore z-axis passing through the two constrictions ($z = 19$ Åand $z = 48$ Å) and the vestibule ($z = 70$ Å) for the four homopeptides.*

The occupancy map $\tilde{M}_{x,y,z}$ allows direct estimation of the pore section $A(z)$ that can be calculated summing $\tilde{M}_{x,y,z}$ on slices of width Δz normal to the pore axis, in formula

$$A_z = \sum_{x,y} \tilde{M}_{x,y,z} \Delta x \Delta y \tag{4.2}$$

Consequently, the resistance, equation 4.1, can be approximated as

$$\tilde{R} = \sum_{i=1}^{N_z} \frac{\rho}{A_z} \Delta z \tag{4.3}$$

where $i = 1$ and $i = N_z$ correspond to the slice at pore trans and vestibule entrances, respectively, and we assumed that the resistivity ρ is constant along the pore. A similar quasi-1D model was recently applied in [97]. Figure 4.3b reports the inverse of the available section profile, A_z^{-1} for the five Ala replicas (solid lines) and for the empty pore (black dashed line). In the vestibule region, $z \in$ (60,100) Å, the difference between the empty and the clogged αHL is negligible, indicating that, the contribution of the moiety of the homopeptide in the vestibule region to the pore resistance, equation 4.3, is almost unrelevant. More evident differences are present in the barrel region, $z \simeq$ (5,50) Å, and, in particular in the main αHL constriction, $z \simeq 50$ Å. Interestingly, $A(z)^{-1}$ has a peak also for $z \simeq 20$ Å. We will indicate this region as secondary barrel constriction. To quantify the correlation between the pore clogging and the ionic current, we defined the pore clogging estimator as

$$b = 1 - \frac{\tilde{R}_0}{\tilde{R}} \tag{4.4}$$

where $\tilde{R}_0$ refers to the empty pore. The above discussed model is based on several hypotheses that are violated by the actual αHL pore shape. In particular, the continuum assumption is not justified at nanoscale, moreover, the model implicitly assumes a smooth variation of $A(z)$ along the pore axis. In addition, we considered a homogeneous electrolyte resistivity ρ. Nevertheless, although a strict quantitative agreement with the blockage $\Delta I/I_0$ and b is not expected, b results to be highly correlated with the measured $\Delta I/I_0$ (Pearson correlation coefficient $r = 0.8$), see Figure 4.4.

The last we looked for a less computational demanding strategy to estimate b. We repeated the protocol described in the section 3.1.2 using, as input data, 32 ns equilibrium runs ($E_z = 0$) instead of the original 240 ns non-equilibrium trajectories. The resulting pore clogging estimator is indicated as b^{eq}. The result show that, although the value of b slightly changes when using equilibrium or non equilibriums runs as input, the correlation is still very good for Ala, Gln and Trp while deviations are obtained for Phe that, at equilibrium, show a larger clogging compared to non-equilibrium runs. The relatively small computational cost of the equilibrium simulations needed to estimate b^{eq} allows us to explore the blockage features of all the amino acids. For each homopeptide, we run five different replicas. Figure 4.5a shows the pore clogging estimator b^{eq} VS the apparent amino acid volume V_a.

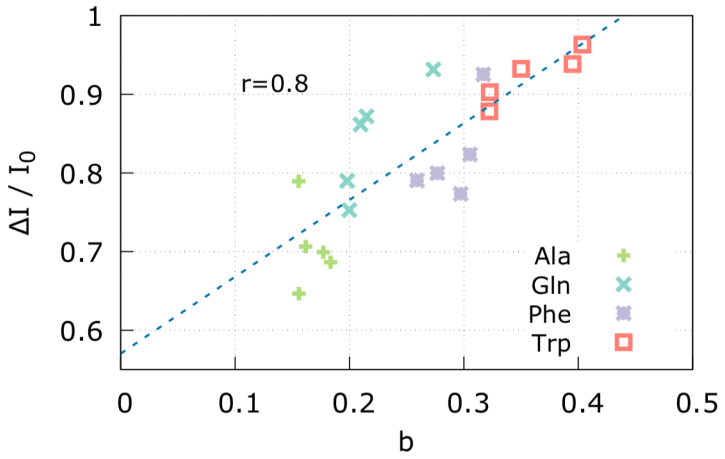


Figure 4.4: *Pore clogging b VS measured current blockage $\Delta I/I_0$. Linear regression curve is reported in dashed blue, Pearson correlation coefficient $r = 0.8$.*

A very good correlation is evident for all uncharged residues, while charged residues lie below the regression line. Indeed, charged residues leave more room to electrolyte solution compared to uncharged one. Similarly, although less evident, polar residues (green) show, on average, a lower b^{eq} than hydrophobic ones b^{eq}. This occurrence can be explained considering that hydrophobic, hydrophilic and charged residues affect the structure of the first shells of the electrolyte solution surrounding them in a different way. Concerning water molecules, hydrophilic and charged residues induce a more compact layering with respect to hydrophobic ones [98]. In addition, charged residues also induced an overall increases of the total number of ions. Hence, on average, for a similar amino acid volume, the pore clogging is minimum for charged residues and it progressively increases moving to hydrophilic and hydrophobic residues. This evidence suggest that, although the main features controlling the pore clogging is the volume of the amino acid, also charge and hydrophobicity play a role. Concerning the uncharged residues, the more evident outlier in Figure 4.5a, among the uncharged residues, is Leu, indeed, although its volume is the same of its isomer Ile, b^{eq} is much larger. A close inspection to the inverse of the available section profile $1/A(z)$ indicates that this discrepancy is mainly due to clogging in the secondary barrel constriction, z $\simeq$ 20 Å. In particular, we observed that in some replicas, the Leu-homopeptide forms a short α-helix in the portion that occupies the secondary constriction. The effect of secondary barrel constriction can be, in principle, eliminated using truncated αHL [9] where it was shown that stable αHL pores are stable also when the large portion of the trans side of the bar-

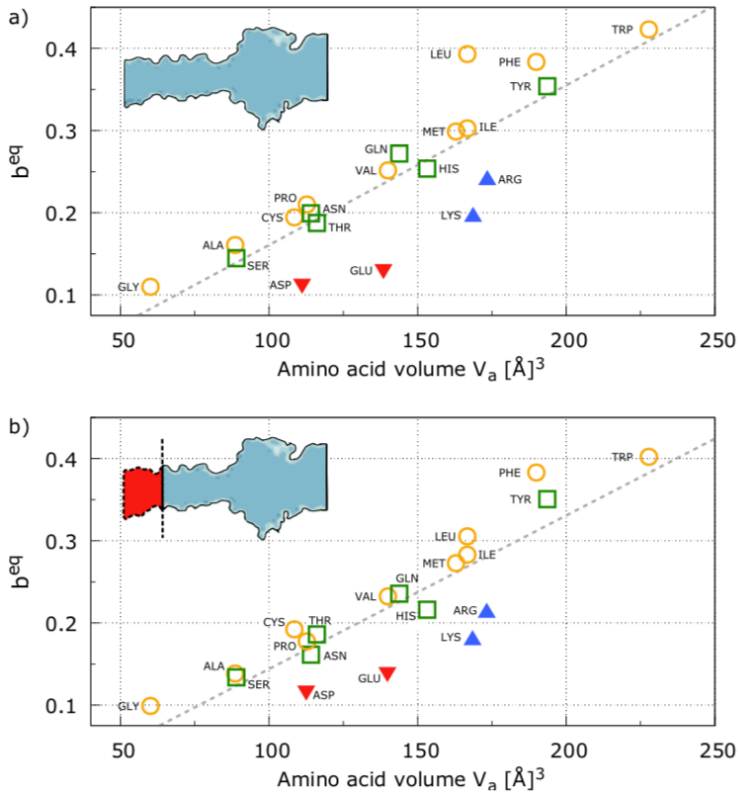


Figure 4.5: *Pore clogging estimator b^{eq} for all residues Vs the amino acid volume V_a. Yellow circles corresponds to hydrophobic residues, green squares to polar, blue up-triangles to positively charged residues and red down-triangles to negatively charged ones. The dashed line is the minimum square fit. Panel a reports the b^{eq} calculated on the entire pore while panel b refers to the b^{eq} calculated removing the last part of the barrel including the secondary constriction (light magenta section in the sketch in the inset).*

rel are deleted. We explored this possibility with our model calculating the summation in equation 4.3 only for the αHL region going from the residues Ile 136 and Asn 123, approximatively 20 from the native trans barrel entrance, to the vestibule. Figure 4.5b reports the corresponding b^{eq} VS V_a plot where Leu lies on the regression line. It is worth noting that recent experiments indicates that αHL is able to distinguish among three-block peptides where the central neutral residues were Alanine and Triptophan [99]. Taken together, the cited experimental results and our current simulations, suggest that biological pore can potentially been employed for protein sequencing although several challenging issues, such as the translocation control, need to be solved [100].

4.2 Elastic Signal

As metioned in the chapter 3.2.2, the current signals have been measured on peptides with applied voltage $\Delta V = 1$ V between the leads. The calculated currents signals have been obtained under quasi-static conditions with an idealized quasi-stationary translocation dynamics where the atoms are allowed to relax at T = 0 K at each translocation stage.

Glycine homo-peptide

Because the Gly side chain is a single hydrogen atom, the amino acid (AA) side chain effects on the peptide bond (PB) current signal should be, in principle, minimized; this circumstance, together with the neutral and non-polar nature of Gly, makes the present one a reference case of the PB signal.

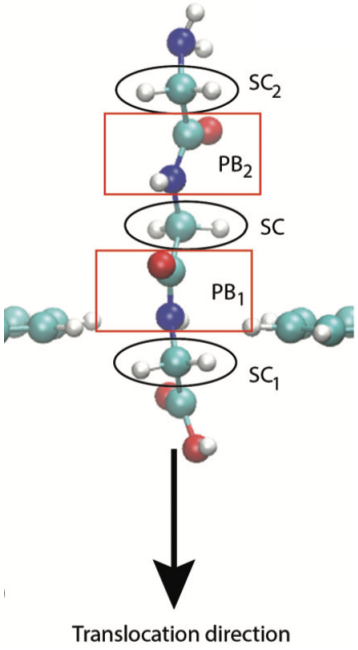


Figure 4.6: *Glycine homo-peptide between nano-gap of graphene*

In Figure 4.7, the current signal across the graphene nano ribbon (GNR) nano-gap is reported as a function of the position of the peptide inside the nano-gap. Each peptide bond appears with two, well-defined, current peaks. The electronic current order of magnitude is comparable with typical tunneling current signals measured in nanogap device but with the advantage of an atomic resolution.

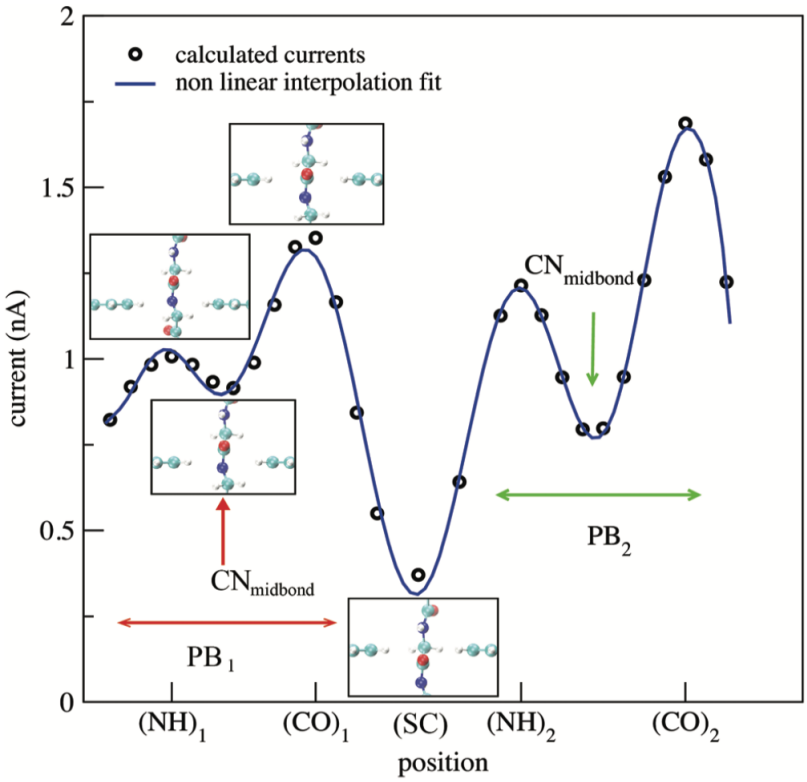


Figure 4.7: *Electron current measured in the ZGNR based nano-gap device while a Gly peptide with three AAs translocates across the nano-gap*

The two peaks correspond approximately to the two positions of, respectively, the N and the O atoms of the peptide bond in the nano-gap and they are separated by a current drop associated with the presence of the C-N bond in the nano-gap. The signal in between two peptide bonds, i.e. with the Gly side chain in the nano-gap, is much smaller. To explain the observed currents, we have analyzed the transmission coefficient functions T(ϵ) corresponding to (CO), (NH) and (CN) configurations. The data are reported in figure 4.8(a) where the continuous (dashed) curves refer to $PB_{1,2}$.

The transmission coefficient curves show that the contribution to the current comes from the states in the energy range [0.2 eV $\leq \epsilon \leq$ 0.5 eV]; moreover, according to the current values of Figure 4.7, for each PB the largest T(ϵ) values are obtained for $(CO)_{1,2}$ configurations (the CO group is in the middle of the ZGNR gap), the second largest occur for $(NH)_{1,2}$ configurations and the smallest ones are obtained with the CN mid-bond in the middle point of the ZGNR gap. For comparison, we reported in Figure 4.8(b) the transmission functions of the empty nano-gap showing that no

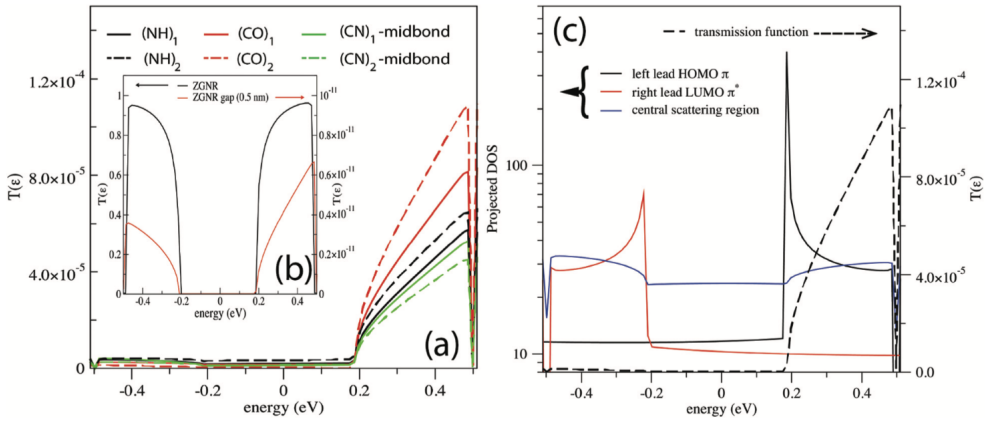


Figure 4.8: *Transmission coefficients at different positions during the translocation of the peptide with three Gly (a); the continuous and dashed curves refer to the first and second PB, respectively. The inset (b) shows the transmission functions of the ZGNR and of the void nano-gap. In (c) is reported the transmission function calculated with the nitrogen atom of the PB in the gap and the PDOS of the left and right leads and of the central scattering region.*

transmission occurs in this case. Moreover, the projected density of states (PDOS), Figure 4.8(c), evidences that the increase in T(ϵ) occurs in correspondence with the increase in the PDOS belonging to the scattering central region and the left-hand side ZGNR HOMO π-like states, provided that the PDOS of the right-hand side ZGNR LUMO π^*-like states is not zero. The channel decomposition of the transmission coefficient calculated at V = 1 V indicates the existence of just one channel contributing to the whole transmission coefficient, independently on the configuration considered with either the CO bond or the NH bond at the center of the GNR gap. To evidence the main contributions to the transmission functions at the current peaks, the atomic PDOS decompositions in the localized basis set of the central scattering region; the main features are reported in Figure 4.9 for the two PBs.

For the $NH_{1,2}$ configurations, the largest density comes from the Gly α-carbon and side chain that are the closest ones at the C-terminal side chain (SC_1 and SC for PB_1 and PB_2). We will indicate these groups as "lower" since they are just below $NH_{1,2}$ in the snapshot in Figure 4.6. Minor contributions to the PDOS of the scattering region come from the PBi itself (with similar $NH_{1,2}$ and $CO_{1,2}$ PDOS values) and the closest side chain group at the N-terminal side (indicated as "upper" side chain in the following), namely SC and SC_2 for PB_1 and PB_2. At the second peak occurring with $CO_{1,2}$ in the nano-gap, the major contribution comes from the "upper" side

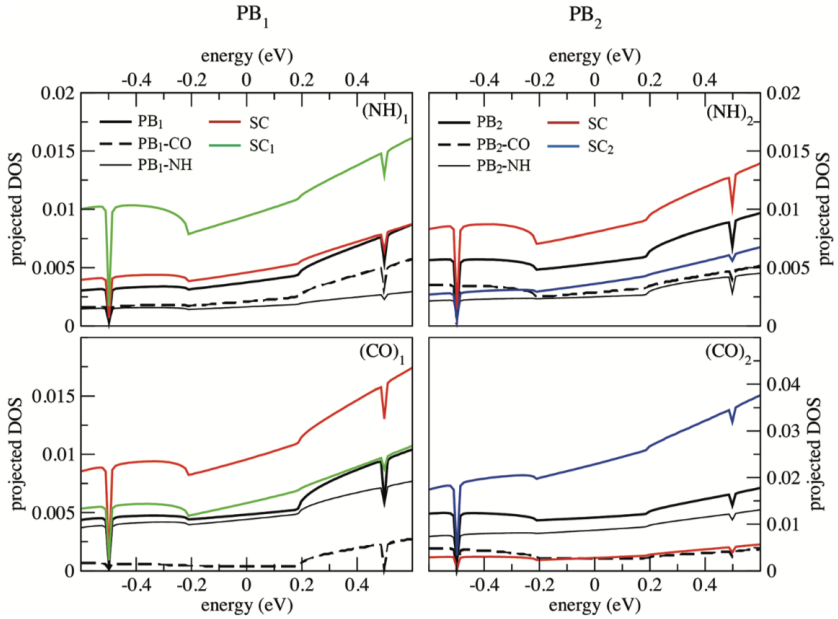


Figure 4.9: *Atomic PDOS decomposition corresponding to the four current maxima measured in the three Gly peptides. The cases with the NH and CO groups in the middle of the nano-gap for each of the two PBs are aligned in columns (with the same legend). The atomic PDOS have been grouped together for atoms belonging to the same unit.*

chain (i.e. the closest one). Concerning the PB-related PDOS, the major contribution comes from the $NH_{1,2}$ group, while the one from the $CO_{1,2}$ group is always lower in the entire energy range. A more detailed analysis of the DOS projected on the PB atomic orbitals evidences the leading role played by the p_z orbitals of N and C. A deeper insight on the local transport phenomena occurring at the atomistic level can be achieved looking at the local currents flowing across the atomic sites (Table 4.1); the resulting scenario is sketched in Figure 4.10 where the most relevant local atomic currents are represented by arrows for $NH_{1,2}$ (a) and $CO_{1,2}$ (b) configurations.

Table 4.1: *Main local atomic current contributions to the whole current at the left lead (electron injection from the left lead); i = 1, 2 indicates the first or second PB. The groups and the corresponding local currents calculated for PB_2 are given in parentheses; bond current corresponds the electron injection into the left lead.*

Configuration	Atomic group	Bond current (nA)
NH_i	$SC_1(SC)$	0.67 (0.857)
	NH_i	0.045 (0.152)
CO_i	$SC(SC_2)$	0.63 (0.773)
	NH_i	0.44 (0.529)
SC	NH_1	0.114
	SC	-0.114
	NH_2	0.146

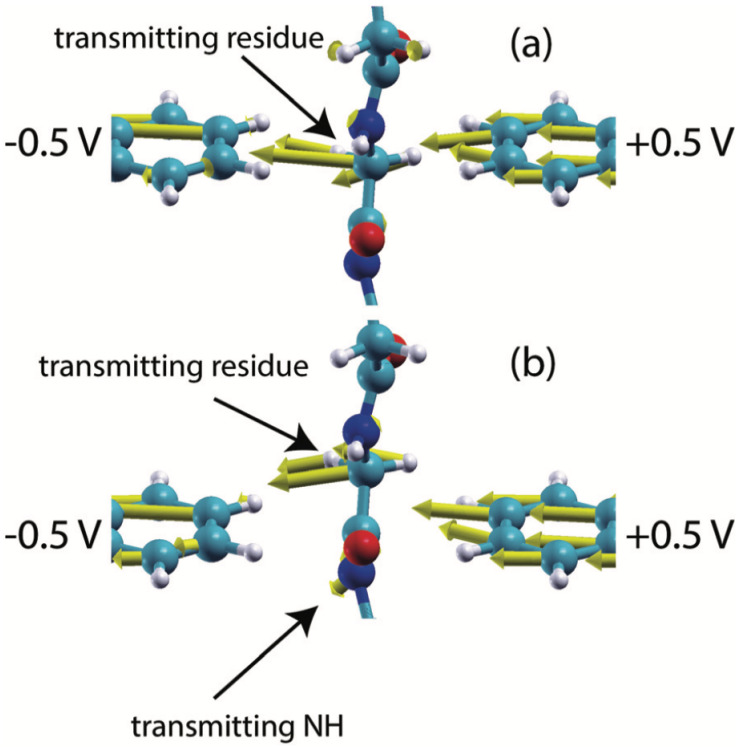


Figure 4.10: *Schematic drawing of the atomic current flows with the NH (a) and CO (b) groups in the middle of the nano-gap. C, O, and N atoms are respectively in cyan, red, and blue and atomic currents are represented by yellow arrows. Transmission occurs via the "upper" and "lower" side chains and in the case of the CO group in the gap (b) also via the nitrogen atom of the PB.*

$(NH)_i$ configurations. Consistent with the corresponding PDOS, the bond current analysis confirms that the main contributions for $(NH)_i$ configurations come from the "lower" side chain groups, i.e. the closest ones at the C-terminal side, namely SC_1 and SC for PB_1 and PB_2, respectively. Indeed, since these groups are off the ZGNR plane but close to it, the overlap of the corresponding atomic orbitals with the left lead π orbital is favored. Much minor contributions come from the upper residues and from the PBs themselves, the last ones, however, behaving in a rather complex way. Indeed a detailed analysis reveals that the CO group acts as reflectors while the NH groups may participate in the current transfer to a very small extent through the N_{p_z} orbital.

$(CO)_i$ configurations. The same analysis of the second current peaks evidences that electrons are injected from the left lead mainly through both the "upper" α-carbon and side chain, i.e. the closest to the CO group at the N-terminal side, and, remarkably, through the NH group of the PB to almost the same extent as the "upper" residue. Then, what makes the difference between the two peaks in a given PB signal is the contribution from the NH group that is rather small in the case of the first peak (NH group in the gap) while it is almost of the same magnitude as the side chain contribution in the case of the second peak. This is mainly due to the off-plane position of the NH group in this second case that favors the overlap and hybridization between the N_{p_z} orbital and the π orbital of the left lead. A complete and detailed analysis of the atomic currents at both the leads reveals, moreover, the occurrence of charge transfer phenomena inside the peptide, the reflection of electrons at the CO sites and minor, but still meaningful, contributions to the current from the second nearest residues.

As a matter of fact, the atomic current analysis reveals that the main reason why the second peak of the PB fingerprint is larger than the first one resides in the larger contribution from the NH group of the PB itself. This behavior is related to the nature of the PB resonating double bond with a partial occupation of the N_{p_z} orbital that, as a consequence, can participate in the electron transfer, provided it is off the ZGNR plane. Finally, we emphasize that the difference between the second and first current peaks in each PB is a clear fingerprint of the PB itself because it is due to the electrons flowing only across the NH group of the PB.

SC configurations. Let's now consider the configuration with the central Gly side chain in the gap where the current signal drops. The corresponding PDOS and schematic drawing of the current fluxes are reported in Figure 4.11.

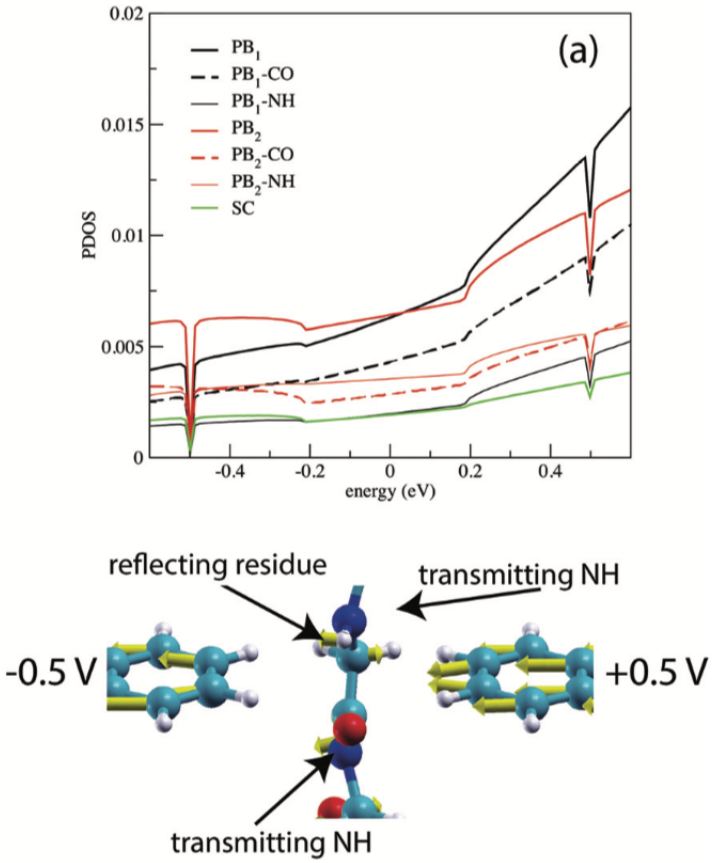


Figure 4.11: *PDOS (a) and schematic drawing of the atomic current flows (b) with the Gly residue in the middle of the nano-gap. C, O, and N atoms are respectively in cyan, red, and blue and atomic currents are represented by yellow arrows. Transmission occurs via the nitrogen atom of the PB while the atomic current is reflected from the central Gly residue.*

Here we see that PB_1 has the largest PDOS followed by PB_2 while the side chain contribution is small in the whole energy range. The main atomic current fluxes at the left lead are reported in Table 4.1. Electrons are injected from the left lead mainly through the closest $(NH)_2$ group, which, however, does not show the largest PDOS (see Figure 4.11(a)), and $(NH)_1$ of PB_1 to nearly the same extent. Contrary to the configurations of the two current peaks, the central side chain SC behaves as a reflecting group. A more de-

tailed analysis shows that a meaningful charge redistribution between the two PBs occurs after injection and also that CO groups may participate in the current flow to a minor, but still meaningful. Therefore, because the Gly residue does not participate in the current transfer, the measured signal is a fingerprint of only the two PBs involved, with a prominent role of the "upper" $(NH)_2$ group. Interestingly, we observe that the total current measured for this configuration is nearly the same as the bond current contributions through the NH group at the second PB peaks ($(CO)_i$ configurations). We can thus conclude the discussion on the Gly homo-peptide emphasizing that for the $(CO)_i$ configurations the electron injection through the PB NH group is maximized, reflecting a large overlap between the partially occupied N_{p_z} orbital and π, it is minimized for $(NH)_i$ configurations with the NH group in the gap, while, in the SC configuration, it separates in two branches across the upper and lower NH groups, with nearly the same total amount as the one measured in the $(CO)_i$ case, thus reflecting a partial overlap between the N_{p_z} and π orbitals.

Hetero-peptides

The other chains of polypeptides analyzed are mixed chains of glycine and alanine (ALA), asparagine (ASN), aspartate (ASP).

The Ala-Gly peptides. The first is a tetra-peptide with alternating Gly and Ala AAs. The tunneling current signal obtained in this case is reported in Figure 4.12. We see that the main features found in the Gly homo-peptide are also present in this case. In particular, the PB current signals are, again, double peaked and the Gly-Ala-Gly transverse current evidences, as expected, a periodic behavior with a doubled period with respect to the Gly-Gly case: indeed the difference between the second and first current maxima, which was nearly the same for PB_1 and PB_2 in the Gly case, now varies between the two adjacent PBs. The signal measured is of the same order of magnitude as in the Gly-Gly case that makes it, again, fully recognizable and separable from the noise due to water molecules. An important feature is that the minimum current values measured, $I \approx 0.4$ nA, are nearly equal to the ones measured in the Gly case and correspond to the Gly and Ala side chains in the middle of the gap. Hence, also in the Gly-Ala case the minimum current signal occurs with the suppression of the transport through the side chain and with the contributions of the upper and lower PBs, with the main role played by the (NH) groups. Looking at the first PB, we see that the current measured for $(CO)_1$ is larger than the one for $(NH)_1$ but their difference is smaller than the minimum current value measured (see the Gly case). This circumstance reflects the difference between the upper (Ala)

and lower (Gly) side chains. The same comments also apply for the second PB (Ala-Gly) where the two peaks are nearly identical.

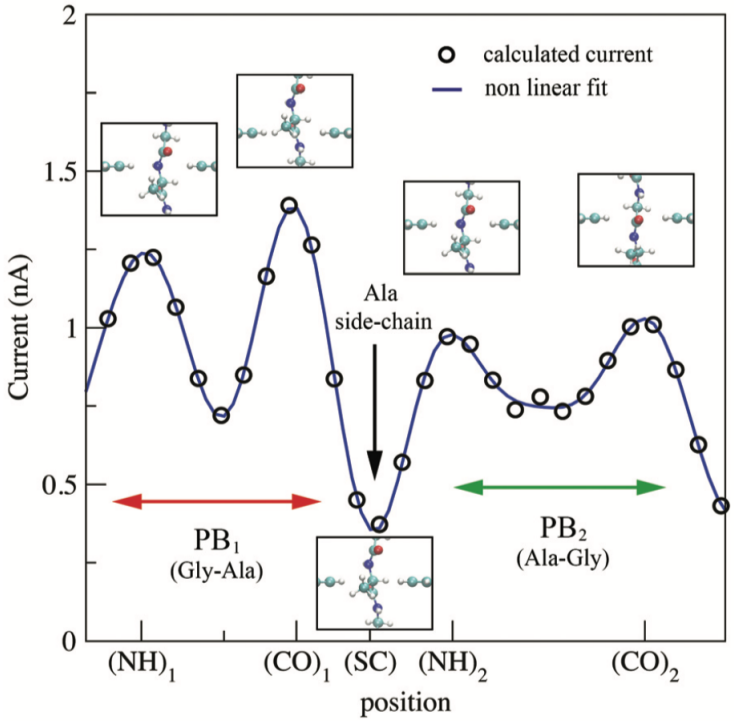


Figure 4.12: *Transverse tunneling current signal from the Gly-Ala peptide. The Gly-Ala and Ala-Gly PBs are considered with the Ala side chain in between.*

The Asn-Gly peptides. The second is the hetero-peptide with alternating Gly and Asn AAs. The tunneling current signal obtained in this case is reported in Figure 4.13. We see also that the main features found in the Gly homo-peptide are in part present in this case. In particular, the PB current signals are single-peaked for Gly-Asn-Gly, this is due to the big difference of peaks of transverse current between the groups that form the peptide-bond (NH peaks are higher than CO peaks), totally the current signal is greater than Gly-Gly homo-peptide. The signal measured is of the same order of magnitude as in the Gly-Gly case that makes it, again, fully recognizable and separable from the noise due to water molecules. An important feature is that the minimum current values measured, I $\approx$ 0.4 nA, are nearly equal to the ones measured in the Gly case and corresponds to the Gly and Asn side chains in the middle of the gap. Also in the Gly-Asn case the minimum current signal occurs with the suppression of the transport through the side chain and with the contributions of the upper and lower PBs, with the main role played by the (NH) groups. we see that the current measured for $(NH)_i$

is larger than the Gly homo-peptide, this circumstance can depend on the polarity of the side chain of Asn.

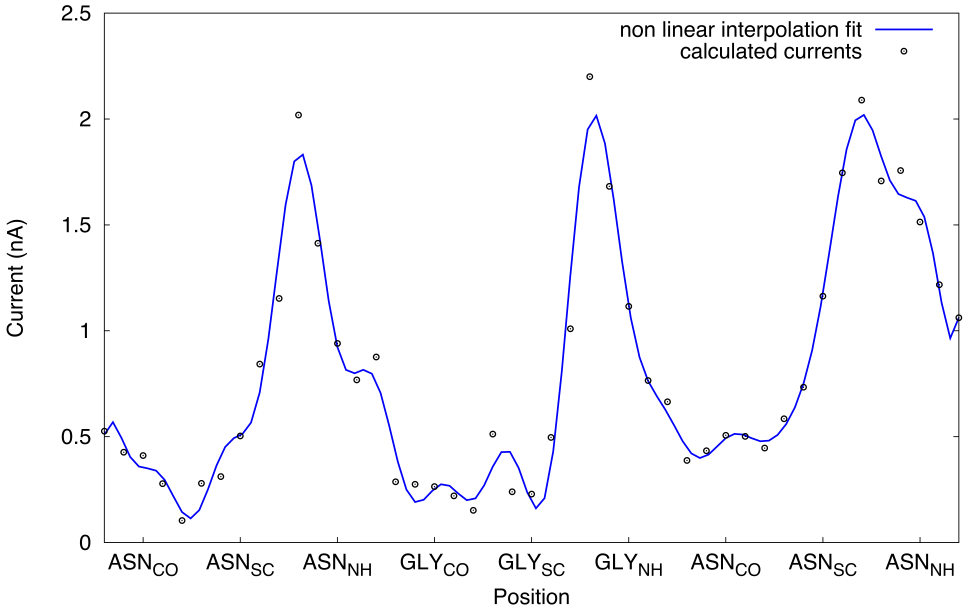


Figure 4.13: *Transverse tunneling current signal from the Gly-Asn peptide.*

The Asp-Gly peptides. The last is the poly-peptide with alternating Gly and Asp AAs. The tunneling current signal obtained in this case is reported in Figure 4.14. We see that the main features found in the Gly homo-peptide are in part present in this case. In particular, the PB current signals are, again, double-peaked, even if less evident, and the Gly-Asp-Gly transverse current evidences a periodic behavior, the signal measured is of the same order of magnitude as in the Gly-Gly case that makes it, again, fully recognizable and separable from the noise due to water molecules. The minimum current value measured, I $\approx$ 0.2 nA, is nearly equal to the ones measured in the Gly case and correspond to the Gly and Asp side chains in the middle of the gap. In the Gly-Asp case, the current signal goes up faster when we move near the side-chain differently from the previous cases. This circumstance reflects the difference between a charged side-chain (Asp) and the non-charged side chains.

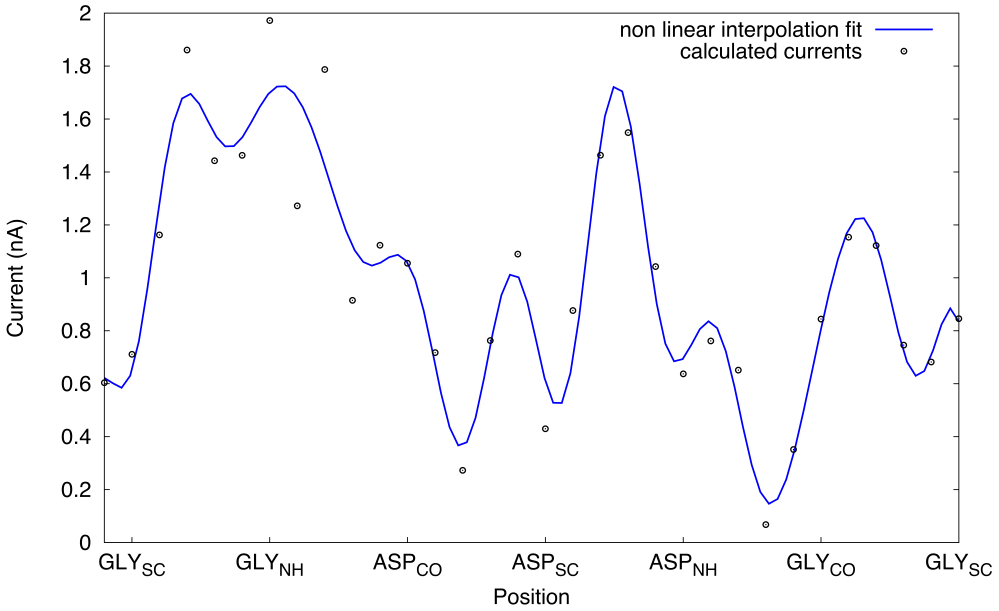


Figure 4.14: *Transverse tunneling current signal from the Gly-Asp peptide.*

Finally, it should be outlined, also, that the narrow size of the GNRs here employed is probably unfit to capture the entire contribution from the hetero-peptides side chain if the peptide backbone is right in the GNR gap center. This enforces our idea of separating the signals from the side chain and the central peptide backbone.

4.3 Inelastic Signal

To better understand what happens inside our nano-gap when the electron current flows across it, we add a last analysis calculating the inelastic signal. We use the main configurations of our peptide of reference, the Gly homo-peptide chain, within the nano-gap of zig-zag graphene nanoribbon (figure 4.6). We use the revised LOE method to calculate the inelastic transmission due to electron-phonon coupling. To correctly calculate the inelastic current, the analysis are carried out near the Fermi energy. Looking at the inelastic current (Figure 4.15) for each configuration, calculated with the formula 2.47, it is possible to see that there are different changes of slope and, except for the SC case, they are similar between the NH and CO configurations.

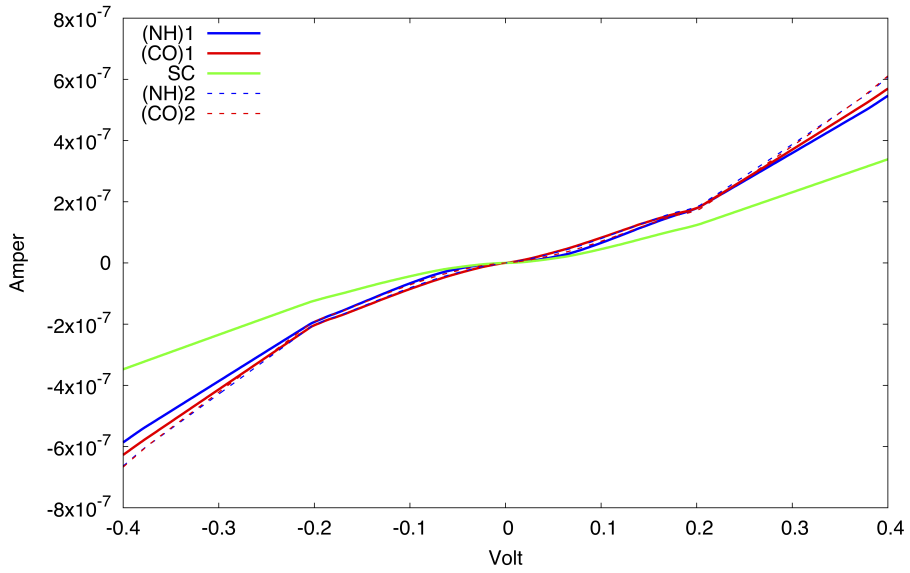


Figure 4.15: *Inelastic current signal for the configurations of the Gly homo-peptide .*

Finally, we show the Phonon-DOS (figure 4.16) and IETS signal (figure 4.17). We can see that the phonon-DOS (figure 4.16) are similar for each group of the Gly homo-peptide, but for IETS signal (figure 4.17) around ≈ 0.37 eV the peaks belong only to the groups that form the peptide-bond ($(CO)_{1,2}$, $(NH)_{1,2}$). This is due to the absence of e-ph coupling of the side-chain, unlike for the CO NH groups. This confirms that the IETS signal can be used to distinguish and count the amino acids of a peptide chain.

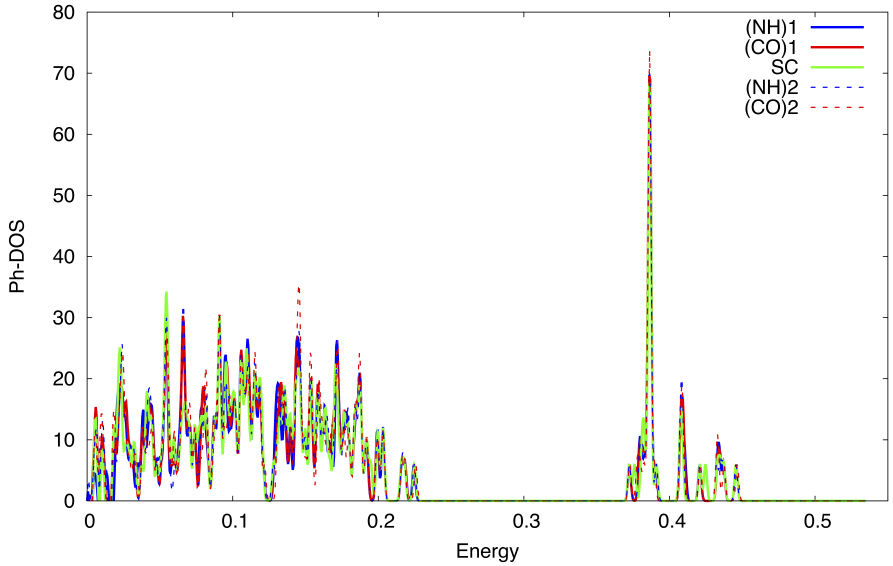


Figure 4.16: *Phonon-DOS for the configurations of the Gly between nano-gap.*

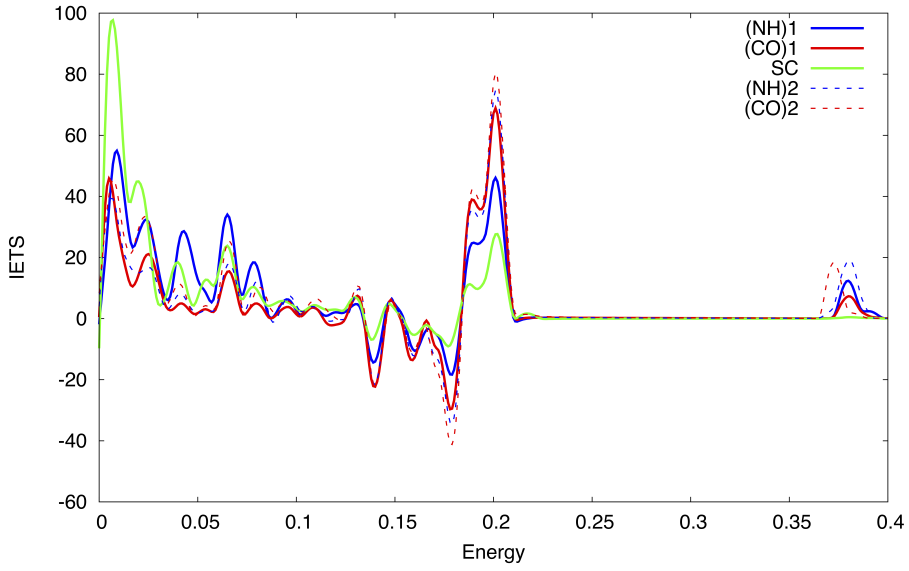


Figure 4.17: *IETS signal, $\frac{d^2I/dV^2}{dI/dV}$, for the configurations of the Gly between nano-gap.*

Chapter 5

Conclusion

We have different conclusions each for the method used, but both indicate how the use of nanopore / nanogap devices for the sequencing of polypetid chains is feasible, indeed the union of the two methods could be the answer for the single-molecule sequencing.

5.1 Classical Molecular Dynamics

We performed an extensive set of all-atom MD simulation to computationally assess the capability of αHL to distinguish among the different amino acids. Inspired by a quasi-1D model for pore conductance, we defined the pore clogging estimator and show that it correlates with the observed current blockage from non-equilibrium runs. Amino acid volume is the main feature that rules the pore clogging and, consequently, the current blockage. Our data clearly show that also hydrophobicity plays a role. Indeed, for similar amino acid volumes, charged residues are associated to a smaller pore clogging than uncharged one and slight, but significant, differences are observed also between hydrophobic and polar amino acid. Our results suggest that αHL is potentially able to discern among the different residues. For some set of residues with very similar volume, however, the pore clogging is very close and the expected current blockage signal as well. In these cases, long current recordings and signal post-processing would be required to distinguish among them.

5.2 Quantum Molecular Dynamics

GNR based nano-devices, made of an array of nano-gaps, have been considered and studied in the context of DFT for single PB detection in a Gly based polypeptide. The peptide translocation across the nano-gap array allows, in principle, the detection of various signals with atomic resolution. Using the NEGF method and the Landauer-B*ü*ttiker formula we show that PBs from Gly polypeptides leave a clear tunneling current fingerprint made of two, well-defined current peaks of the order of magnitude of nA. We have shown that this signal, a sort of a reference PB signal, arises from one channeled transmission function due to the superposition of the left lead pseudo-π HOMO, the right lead pseudo-π^* LUMO and the α-carbons, side chains and PB related orbitals. The extent of these different contributions depends on the relative positions of PBs and residues with respect to the GNR plane. The peaks of the current signal, occurring with the NH and CO groups of the PB in the middle of the GNR nano-gap, are mainly due to the overlap of the closest off-plane residues and the π and π^* GNR states. Interestingly, the difference between the peaks (CO) and the peaks (NH) is due the atomic current flowing across the NH group that contributes to the total current through the off-plane overlap between N_{p_z} and the HOMO and LUMO GNR states. The PB contribution to the current is the only one that survives when the side chain residue is in the middle of the nano-gap giving the same current level for each polypeptide. A prominent role is played by the Npz orbital but a complex scenario, involving also charge transfer across the atoms of the PB, emerges when looking at all the orbitals involved. The described general behavior is confirmed when we considered other peptides alternating Gly and non-polar, polar and charge residues; that still shows peaks for current signal for each PB measured. The reported results pave the way for the further development of nano-devices made of graphene nano-gaps for AA detection and protein sequencing [101].

Appendix A

Appendices

In this chapter, we describe the main simulation tools used for the realization of this thesis work.

A.1 NAMD

NAMD (NAnoscale Molecular Dynamics) is a parallel molecular dynamics code designed for high-performance simulation of large bio-molecular systems, that performs atomic simulations in which the atoms move according to the Newtonian equations of motion. The computational techniques only provide the ability to approximate the solutions. NAMD uses an all-atom MD simulation, which assumes that, every atom experiences a model force field, which accounts for the interaction between an individual atom and all the other atoms in the simulations. The Newtonian equations of motion are modified so that the computed short time step can still be interpreted correctly. In this way, NAMD uses a stochastic coupling approach to enhance the dynamic stability, the stochastic Langevin equation is used in NAMD to generate the Boltzmann distribution, a probability measure for the distribution of the states of a system, for canonical ensemble simulations. The generic Langevin equation is:

$$M\dot{v} = F - \gamma v + \sqrt{\frac{2\gamma k_B T}{M}} R \tag{A.1}$$

Where M is the mass, v is the velocity, F is the force, r is the position, γ is the friction coefficient, k_B is the Boltzmann constant, T is the temperature and R is a univariate Gaussian random process.This equation governs the simulation interaction between the atoms in the simulated molecule. NAMD scales to hundreds of processors on high-end parallel platforms, as well as tens

of processors on low-cost commodity clusters, and also runs on individual desktop and laptop computers. NAMD works with AMBER and CHARMM potential functions, parameters, and file formats [64].

A.2 VMD

VMD (Visual Molecular Dynamics) is a molecular graphics program designed for the display and analysis of molecular assemblies, in particular biopolymers such as proteins and nucleic acids. VMD can simultaneously display any number of structures using a wide variety of rendering styles and coloring methods. Molecules are displayed as one or more representations, in which each representation embodies a particular rendering method and coloring scheme for a selected subset of atoms. The atoms displayed in each representation are chosen using an extensive atom selection syntax, which includes Boolean operators and regular expressions. VMD provides a complete graphical user interface for program control, as well as a text interface using the Tcl embeddable parser to allow for complex scripts with variable substitution, control loops, and function calls. Full session logging is supported, which produces a VMD command script for later playback. High-resolution raster images of displayed molecules may be produced by generating input scripts for use by a number of photorealistic image-rendering applications. VMD has also been expressly designed with the ability to animate molecular dynamics (MD) simulation trajectories, imported either from files or from a direct connection to a running MD simulation. VMD is the visualization component of MDScope, a set of tools for interactive problem solving in structural biology, which also includes the parallel MD program NAMD, and the MDCOMM software used to connect the visualization and simulation programs, VMD is written in C++ [70].

A.3 CHARMM

CHARMM (Chemistry at HARvard Molecular Mechanics) [65] is a highly versatile and widely used molecular simulation program. It has been developed over the last three decades with a primary focus on molecules of biological interest, including proteins, peptides, lipids, nucleic acids, carbohydrates, and small molecule ligands, as they occur in solution, crystals, and membrane environments. For the study of such systems, the program provides a large suite of computational tools that include numerous conformational and path sampling methods, free energy estimators, molecular minimization, dynamics, and analysis techniques, and model-building capabilities. The CHARMM program is applicable to problems involving a much broader class of many-particle systems. Calculations with CHARMM can be performed using a number of different energy functions and models, from mixed quantum mechanical-molecular mechanical force fields, to all-atom classical potential energy functions with explicit solvent and various boundary conditions, to implicit solvent and membrane models. The program has been ported to numerous platforms in both serial and parallel architectures. The general form of the potential energy function most commonly used in CHARMM for macromolecular simulations is based on fixed point charges is:

$$
\begin{aligned}
U(\vec{R}) &= \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 \\
&+ \sum_{Urey-Bradey} K_{UB}(S-S_o)^2 + \sum_{diedrals} K_\phi(\phi - cos(n\phi-\delta)) + \sum_{impropers} K_\omega(\omega-\omega_0)^2 \\
&+ \sum_{non-bonded} (\epsilon_{ij}^{min}((\frac{R_{ij}^{min}}{r_{ij}})^{12} - 2(\frac{R_{ij}^{min}}{r_{ij}})^6) + \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}}) + \sum_{residues} U_{CMAP}(\phi,\psi)
\end{aligned}
\tag{A.2}
$$

The potential energy $U(\vec{R})$ is a sum over individual terms representing the internal and non-bonded contributions as a function of the atomic coordinates. Internal terms include bond(b), valence angle θ, Urey-Bradley (UB,S), dihedral angle (ϕ), improper angle (ω), and backbone torsional correction ($CMAP$,ϕ,ϕ) contributions, the parameters K_b, K_ϕ, K_{UB}, K_θ and K_ω are the respective force constants and the variables with the subscript 0 are the respective equilibrium values. All the internal terms are taken to be harmonic, except the dihedral angle term, which is a sinusoidal expression; here n is the multiplicity or periodicity of the dihedral angle and δ is the phase

shift. The all-atom implementations of the CHARMM force field include all possible valence and dihedral angles for bonded atoms, and the dihedral angle term about a given bond may be expanded in a Fourier series of up to six terms. Most commonly, one dihedral angle term is used, though two or more have been introduced in some cases. In addition, for the protein main chain a numerical correction term, called CMAP, has been implemented. For three bonded atoms A-B-C, the Urey-Bradley term is a quadratic function of the distance, S, between atoms A and C. The improper dihedral angle term is used at branch points; that is, for atoms A, B, and D bonded to a central atom, C, the term is a quadratic function of the (pseudo)dihedral angle defined by A-B-C-D. Both the Urey-Bradley and improper dihedral terms are used to optimize the fit to vibrational spectra and out-of-plane motions. In the polar hydrogen models, the improper dihedral angle term is also required to prevent inversion of chirality. While the improper dihedral term is used very generally in the CHARMM force fields, the Urey-Bradley term tends to be used only in special cases. Non-bonded terms include Coulombic interactions between the point charges (q_i and q_j) and the Lennard-Jones (LJ) 6-12 term, which is used for the treatment of the core-core repulsion and the attractive van der Waals dispersion interaction. Non-bonded interactions are calculated between all atom pairs within a user-specified interatomic cutoff distance, except for covalently bonded atom pairs (1,2 interactions) and atom pairs separated by two covalent bonds (1,3 interactions). The relative dielectric constant, ?, is set to one in calculations with explicit solvent, corresponding to the permittivity of vacuum, ϵ_0. In addition, the electrostatic term can be scaled using other values for the dielectric constant or a distance-dependent dielectric; in the latter, the electrostatic term is inversely proportional to the distance between the interacting atoms squared. In the LJ term, the well depth is represented by ϵ_{ij}^{min}, where i and j are the indices of the interacting atoms, r_{ij} is the interatomic distance, and R_{ij}^{min} is the distance at which the LJ term has its minimum. Typically, and are obtained for individual atom types and then combined to yield and for the interacting atoms via a standard combination rule. In the current CHARMM force fields the $\epsilon_{ij}min$ values are obtained via the geometric mean ($\epsilon_{ij}^{min} = \sqrt{\epsilon_{ij}^{min}\epsilon_{ij}^{min}}$) and R_{ij}^{min} via the arithmetic mean, $R_{ij}^{min} = (R_{ij}^{min} + R_{ij}^{min})/2$. Separate LJ parameters and a scaling factor for electrostatics can be used for the non-bonded interactions between atoms separated by three covalent bonds (1,4 interactions). The Buckingham potential has recently been added as an alternative to the simple LJ for treating the core repulsion. The Morse potential, often used for bond-breaking, is also implemented. The simple form for the potential energy used in this is represents a compromise between accuracy

and speed. For biomolecules at or near room temperature, the harmonic representation is generally adequate, though approximate, and the same holds true for the use of the Lennard-Jones potential for the van der Waals interactions.

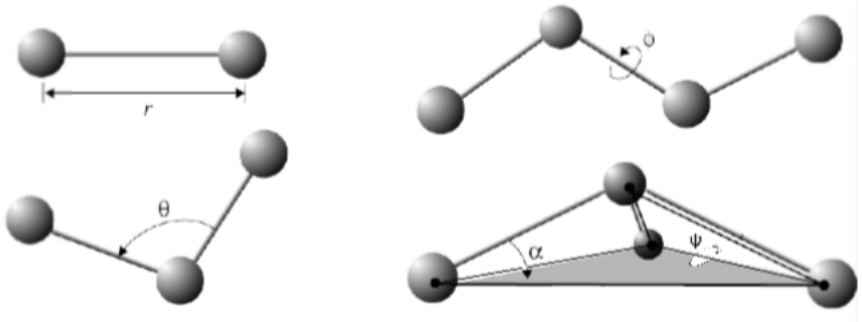


Figure A.1: *r governs bond stretching; θ represents the bond angle term; φ gives the dihedral angle; the small out-of-plane angle α is governed by the so-called ?improper? dihedral angle ψ.*

A.4 SIESTA

SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) is one of the large variety of DFT codes that solves the Kohn-Sham equations; as in many other DFT codes, one uses periodic BCs corresponding to a supercell with periodicity in all three dimensions. This is convenient for treating infinite systems such as crystals, but can also handle finite systems by making the supercell sufficiently large separating the objects. The framework in the supercell approach is Bloch's theorem, which states that for a periodic system the electronic wavefunction can be written as a product of a wavelike part and a cell-periodic part, i.e.,

$$\psi_{j,k(r)} = e^{ikr} u_{j,k(r)} \tag{A.3}$$

where j is a discrete band index and k a reciprocal lattice vector belonging to the first Brillouin zone (BZ) corresponding to the supercell. The theorem allows for mapping the KS eigenvalue problem into the reciprocal space, where one can separately obtain for each k-point a discrete set of eigenstates of the Hamiltonian. The expectation value of some one-body operator $\hat{O}$ is then calculated as

$$\langle \hat{O} \rangle = \frac{1}{\Omega_{BZ}} \int_{BZ} dk O(k) \approx \sum_{k \epsilon BZ} w_k O(k) \tag{A.4}$$

where the integral over the first Brillouin zone (BZ), with volume Ω_{BZ}, for practical purposes is approximated by a sum over k-points with weight factors w_k (adding up to one). In SIESTA this discrete BZ sampling is based on the so-called Monkhorst-Pack [76]. Note at this point that the larger the supercell the smaller the corresponding BZ. For sufficiently large supercells this BZ sampling becomes less critical and using only the Γ-point might be a reasonable approximation. In most DFT implementations one uses pseudopotentials to get rid of the core electrons. The idea is to replace the true atomic potential and the chemically inert core electrons with an effective potential (the pseudopotential) that provides the same description for the valence electrons. As a result the computations simplify since one just has to solve for the valence electronic structure. In SIESTA one generally uses norm-conserving pseudopotentials according to the Troullier-Martins parameterization [77]. For a numerical solution of the Kohn-Sham equations one typically chooses a finite basis by which to represent the wave functions. In SIESTA one uses atomic-like localized orbitals that guarantee the Hamiltonian and overlap matrix to be sparse. For each atom I positioned at RI one defines a set of atom centered orbitals

$$\phi_{I,lmn}(r) = \phi_{I,ln}(r_l) Y_{lmn}(\hat{r}_I) \tag{A.5}$$

where $\phi_{I,ln}(r_l)$ and $Y_{lmn}(\hat{r}_I)$ are radial and angular components, respectively. Distances are conveniently written in terms of $r_l = r - R_I$ and the angular momentum is labeled by l, m. With a multiple-ζ basis there will be several orbitals (labeled n) corresponding to the same angular momentum but with different radial dependence. The basis orbitals are strictly confined in the sense that they are zero beyond a certain radius (which may be different for each of the radial functions). This cutoff radius is usually specified indirectly in terms of a confinement energy. The SIESTA basis implies that the calculation of the overlap matrix and most matrix elements of the Hamiltonian are two-center integrals, which are effectively calculated within Fourier space where the convolution becomes a simple product. However, the density and some Hamiltonian matrix elements are calculated on a real-space grid. The fineness of this grid is conveniently described by an energy "grid cutoff". For a complete description the reader is referred to Ref. [73, 102].

A.5 TranSIESTA

TranSIESTA (TRANsport Spanish Initiative for Electronic Simulations with Thousands of Atoms), enables modelling of molecular electronic devices under operation conditions [74]. The method is based on a procedure to solve the electronic structure of an open system formed by a finite structure sandwiched between two semi-infinite metallic leads. A finite bias can be applied between both leads, to drive a finite current. It is based on the non-equilibrium Green's function (NEGF) technique that has been interfaced with the SIESTA electronic structure package [73] in such a way that the density matrix of the system is calculated self-consistently when the system is subject to an external bias. In practical terms, calculations using TranSIESTA involve the solution of the electronic density from the DFT Hamiltonian using Green's functions techniques, instead of the usual diagonalization procedure. Therefore, TranSIESTA calculations involve a SIESTA run, in which a set of routines are invoked to solve the Green's functions and the charge density for the open system.

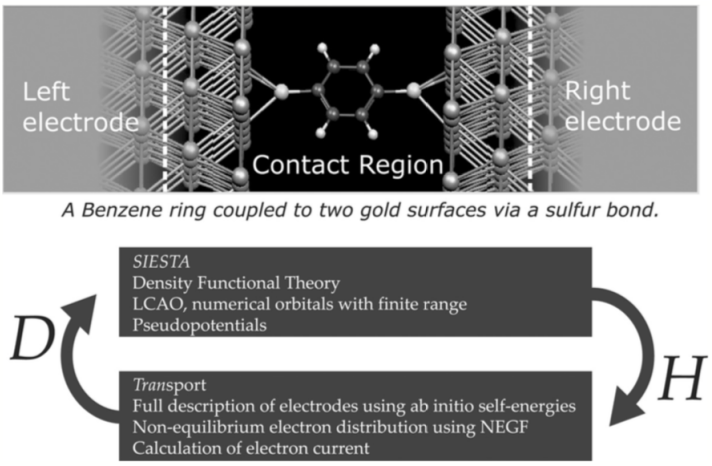


Figure A.2: *A molecule sandwiched between two semi-infinite gold crystals with different electrochemical potentials. The TranSIESTA package uses the SIESTA program to obtain the Hamiltonian, H, in a localized basis set. The Hamiltonian is separated into left electrode, contact region and right electrode. The electrode Hamiltonians are obtained from separate bulk calculations, whereas the Hamiltonian of the contact region is calculated self-consistently. The trans part calculates the non-equilibrium density matrix, D, from H, using a non-equilibrium Green's function (NEGF) technique.*

Figure A.2 shows a typical molecular electronics system, a molecule coupled to two electrodes with different electrochemical potentials. A few layers

into the electrodes the effect of the interaction region is screened away and the atoms are in a bulk environment, thus only a finite region of the system needs to be modelled self-consistently. We separate the system into three parts, left (L), contact (C), and right (R). The Hamiltonians of L and R are obtained from a separate bulk calculations, whereas the Hamiltonian of C is calculated self-consistently. To obtain the density matrix of C we employ a NEGF technique. The parts of the left and right electrode that have matrix elements within the region C we call I_L and I_R, respectively. The coupling of I_L and I_R with the remaining part of the infinite electrodes is taken into account using self-energies, Σ_L and Σ_R. In the direction parallel to the interface we impose periodic boundary conditions. The retarded Green's function, G, is obtained from matrix inversion:

$$G(E) = [(E)S - H - \Sigma_L(E)] - \Sigma_R(E)]^{-1} \tag{A.6}$$

where H is the Hamiltonian and S the overlap matrix of region I_L, C, and I_R. Note that the Green's function calculated from *Equation* A.6 is equivalent to the Green's function obtained by inverting [ES - H] of the infinite system. The non-equilibrium density matrix, D, is related to the retarded Green's function by the equation

$$D = \frac{1}{\pi}\int dE G(E) Im \sum_{L(E)} G^{+}(E) n_F(E-\mu_L) + G(E) Im \sum_{R(E)} G^{+}(E) n_F(E-\mu_R) \tag{A.7}$$

The density functional theory to obtain the Hamiltonian,

$$H_{ij} = \langle i \left| \frac{\hbar}{2m}\nabla^2 + V^{eff}(\rho) \right| j \rangle \tag{A.8}$$

where $|i\rangle$. and $|j\rangle$.are the localized basis functions used by SIESTA. The SIESTA basis sets [103]. are solutions to the Schrödinger equation of the atom with the boundary condition that they should be zero at a specific radius. The implementation includes the possibility of multiple zetas and polarization terms, and the accuracy is similar to calculations with Gaussian orbitals. The atom cores are described by Troullier-Martins pseudopotentials. The self-consistent density matrix of the system is obtained by solving the above equations iteratively using a Pulay mixing scheme [104]. From the self-consistent density matrix we obtain the total energy and forces [105]. The current and transmission coefficients are obtained from the retarded Green's function and the left and right self-energies.

A.6 INELASTICA

INELASTICA is both the name of this whole Python package as well as that of an included script to compute inelastic transport characteristics. INELASTICA is based on the SIESTA/TranSIESTA DFT codes. The project was initiated around 2003-2005 by Thomas Frederiksen and Magnus Paulsson while they worked in the group of Mads Brandbyge at the Technical University of Denmark [94]. INELASTICA contains a number of scripts such as:

- **geom2geom:** Geometry conversion between different file formats
- **Bandstructures:** Computation of electron and phonon band structures
- **pyTBT:** A Python version of tbtrans for elastic electron transport
- **EigenChannels:** Eigenchannel analysis and generation of real-space scattering state wave functions
- **Phonons:** Vibration modes/frequencies and electron-vibration couplings
- **Inelastica:** Inelastic transport characteristics (IETS spectroscopy, inelastic shot noise, local heating, etc.)
- **STM:** Calculation of STM images using the Bardeen approximation

A.7 Quantum ESPRESSO

QUANTUM ESPRESSO [85] (opEn-Source Package for Research in Electronic Structure, Simulation, and Optimization) implements a variety of methods and algorithms aimed at a chemically realistic modeling of materials from the nanoscale upwards, based on the solution of the density functional theory (DFT) problem, using a plane waves (PWs) basis set and pseudopotentials (PPs) to represent electron-ion interactions. The codes are constructed around the use of periodic boundary conditions, which allows for a straightforward treatment of infinite crystalline systems, and an efficient convergence to the thermodynamic limit for aperiodic but extended systems, such as liquids or amorphous materials. QUANTUM ESPRESSO can be used for any crystal structure or supercell, and for metals as well as for insulators. The atomic cores can be described by separable norm-conserving (NC) PPs, ultrasoft (US) PPs, or by projector-augmented wave (PAW) sets. Different exchange-correlation functionals are available in the framework of the local-density (LDA) or generalized-gradient approximation (GGA). The basic computations/simulations that can be performed include:

- calculation of the Kohn-Sham (KS) orbitals and energies for isolated or extended/periodic systems, and of their ground-state energies;
- complete structural optimizations of the microscopic (atomic coordinates) and macroscopic (unit cell) degrees of freedom, using Hellmann-Feynman forces and stresses:
- ab-initio molecular dynamics (MD), using either the Car-Parrinello Lagrangian or the Hellmann-Feynman forces calculated on the Born-Oppenheimer (BO) surface, in a variety of thermodynamical ensembles, including NPT variable-cell MD

Bibliography

[1] Frederick Sanger, Steven Nicklen, and Alan R Coulson. Dna sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12):5463–5467, 1977.

[2] Daniel Branton, David W Deamer, Andre Marziali, Hagan Bayley, Steven A Benner, Thomas Butler, Massimiliano Di Ventra, Slaven Garaj, Andrew Hibbs, Xiaohua Huang, et al. The potential and challenges of nanopore sequencing. In *Nanoscience And Technology: A Collection of Reviews from Nature Journals*, pages 261–268. World Scientific, 2010.

[3] Claudia A Madampage, Omid Tavassoly, Chris Christensen, Meena Kumari, and Jeremy S Lee. Nanopore analysis: An emerging technique for studying the folding and misfolding of proteins. *Prion*, 6(2):116–123, 2012.

[4] Christian B Rosen, David Rodriguez-Larrea, and Hagan Bayley. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nature biotechnology*, 32(2):179, 2014.

[5] Wallace H Coulter. Means for counting particles suspended in a fluid, October 20 1953. US Patent 2,656,508.

[6] James Clarke, Hai-Chen Wu, Lakmal Jayasinghe, Alpesh Patel, Stuart Reid, and Hagan Bayley. Continuous base identification for single-molecule nanopore dna sequencing. *Nature nanotechnology*, 4(4):265, 2009.

[7] David Stoddart, Giovanni Maglia, Ellina Mikhailova, Andrew J Heron, and Hagan Bayley. Multiple base-recognition sites in a biological nanopore: Two heads are better than one. *Angewandte Chemie International Edition*, 49(3):556–559, 2010.

[8] David Stoddart, Andrew J Heron, Ellina Mikhailova, Giovanni Maglia, and Hagan Bayley. Single-nucleotide discrimination in immobilized dna oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences*, 106(19):7702–7707, 2009.

[9] David Stoddart, Mariam Ayub, Lajos Höfler, Pinky Raychaudhuri, Jochen W Klingelhoefer, Giovanni Maglia, Andrew Heron, and Hagan Bayley. Functional truncated membrane pores. *Proceedings of the National Academy of Sciences*, 111(7):2425–2430, 2014.

[10] Jiali Li, Derek Stein, Ciaran McMullan, Daniel Branton, Michael J Aziz, and Jene A Golovchenko. Ion-beam sculpting at nanometre length scales. *Nature*, 412(6843):166, 2001.

[11] AJ Storm, JH Chen, XS Ling, HW Zandbergen, and C Dekker. Fabrication of solid-state nanopores with single-nanometre precision. *Nature materials*, 2(8):537, 2003.

[12] Jiunn B Heng, Chuen Ho, Taekyung Kim, Rolf Timp, Aleksij Aksimentiev, Yelena V Grinkova, Stephen Sligar, Klaus Schulten, and Gregory Timp. Sizing dna using a nanometer-diameter pore. *Biophysical journal*, 87(4):2905–2911, 2004.

[13] Eamonn Kennedy, Zhuxin Dong, Clare Tennant, and Gregory Timp. Reading the primary structure of a protein with 0.07 nm 3 resolution using a subnanometre-diameter pore. *Nature nanotechnology*, 11(11):968, 2016.

[14] Samir M Iqbal, Demir Akin, and Rashid Bashir. Solid-state nanopore channels with dna selectivity. *Nature nanotechnology*, 2(4):243, 2007.

[15] Slaven Garaj, W Hubbard, A Reina, J Kong, D Branton, and JA Golovchenko. Graphene as a subnanometre trans-electrode membrane. *Nature*, 467(7312):190, 2010.

[16] Song Liu, Bo Lu, Qing Zhao, Ji Li, Teng Gao, Yubin Chen, Yanfeng Zhang, Zhongfan Liu, Zhongchao Fan, Fuhua Yang, et al. Boron nitride nanopores: highly sensitive dna single-molecule detectors. *Advanced materials*, 25(33):4549–4554, 2013.

[17] Ke Liu, Jiandong Feng, Andras Kis, and Aleksandra Radenovic. Atomically thin molybdenum disulfide nanopores with high sensitivity for dna translocation. *ACS nano*, 8(3):2504–2511, 2014.

[18] Christopher A Merchant, Ken Healy, Meni Wanunu, Vishva Ray, Neil Peterman, John Bartel, Michael D Fischbein, Kimberly Venta, Zhengtang Luo, AT Charlie Johnson, et al. Dna translocation through graphene nanopores. *Nano letters*, 10(8):2915–2921, 2010.

[19] Grégory F Schneider, Stefan W Kowalczyk, Victor E Calado, Grégory Pandraud, Henny W Zandbergen, Lieven MK Vandersypen, and Cees Dekker. Dna translocation through graphene nanopores. *Nano letters*, 10(8):3163–3167, 2010.

[20] Bo Song, Grégory F Schneider, Qiang Xu, Grégory Pandraud, Cees Dekker, and Henny Zandbergen. Atomic-scale electron-beam sculpting of near-defect-free graphene nanostructures. *Nano letters*, 11(6):2247–2250, 2011.

[21] Jirí Šponer and Filip Lankaš. *Computational Studies of RNA and DNA*, volume 2. Springer Science & Business Media, 2006.

[22] Jiří Šponer, Jerzy Leszczynski, and Pavel Hobza. Electronic properties, hydrogen bonding, stacking, and cation binding of dna and rna bases. *Biopolymers: Original Research on Biomolecules*, 61(1):3–31, 2001.

[23] Engelbert Buxbaum. *Fundamentals of protein structure and function*, volume 31. Springer, 2007.

[24] RY Yada, RL Jackman, and S Nakai. Secondary structure prediction and determination of proteins?a review. *International journal of peptide and protein research*, 31(1):98–108, 1988.

[25] Elaine R Mardis. Next-generation dna sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 9:387–402, 2008.

[26] David A Wheeler, Maithreyan Srinivasan, Michael Egholm, Yufeng Shen, Lei Chen, Amy McGuire, Wen He, Yi-Ju Chen, Vinod Makhijani, G Thomas Roth, et al. The complete genome of an individual by massively parallel dna sequencing. *nature*, 452(7189):872, 2008.

[27] Alexander S Mikheyev and Mandy MY Tin. A first look at the oxford nanopore minion sequencer. *Molecular ecology resources*, 14(6):1097–1102, 2014.

[28] SJ Singer, Pamela A Maher, and Michael P Yaffe. On the translocation of proteins across membranes. *Proceedings of the National Academy of Sciences*, 84(4):1015–1019, 1987.

[29] Qitao Zhao, Dilani A Jayawardhana, Deqiang Wang, and Xiyun Guan. Study of peptide transport through engineered protein channels. *The Journal of Physical Chemistry B*, 113(11):3572–3578, 2009.

[30] Robert Bikwemu, Aaron J Wolfe, Xiangjun Xing, and Liviu Movileanu. Facilitated translocation of polypeptides through a single nanopore. *Journal of Physics: Condensed Matter*, 22(45):454117, 2010.

[31] Eric M Nicholson and J Martin Scholtz. Conformational stability of the escherichia coli hpr protein: test of the linear extrapolation method and a thermodynamic characterization of cold denaturation. *Biochemistry*, 35(35):11369–11378, 1996.

[32] Jim Schnabel. Protein folding: The dark side of proteins. *Nature News*, 464(7290):828–829, 2010.

[33] Ju-Hyun Lee, W Haung Yu, Asok Kumar, Sooyeon Lee, Panaiyur S Mohan, Corrinne M Peterhoff, Devin M Wolfe, Marta Martinez-Vicente, Ashish C Massey, Guy Sovak, et al. Lysosomal proteolysis and autophagy require presenilin 1 and are disrupted by alzheimer-related ps1 mutations. *Cell*, 141(7):1146–1158, 2010.

[34] Nicholas AW Bell and Ulrich F Keyser. Specific protein detection using designed dna carriers and nanopores. *Journal of the American Chemical Society*, 137(5):2035–2041, 2015.

[35] David Rodriguez-Larrea and Hagan Bayley. Multistep protein unfolding during nanopore translocation. *Nature nanotechnology*, 8(4):288, 2013.

[36] Daan Frenkel and Berend Smit. Understanding molecular simulations: from algorithms to applications. Technical report, Academic press, 2002.

[37] Peter L Freddolino, Feng Liu, Martin Gruebele, and Klaus Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding ww domain. *Biophysical journal*, 94(10):L75–L77, 2008.

[38] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[39] Dennis C Rapaport and Dennis C Rapaport Rapaport. *The art of molecular dynamics simulation*. Cambridge university press, 2004.

[40] Alexander D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616, 1998.

[41] Fatemeh Khalili-Araghi, James Gumbart, Po-Chao Wen, Marcos Sotomayor, Emad Tajkhorshid, and Klaus Schulten. Molecular dynamics simulations of membrane channels and transporters. *Current opinion in structural biology*, 19(2):128–137, 2009.

[42] Wei Jiang, David J Hardy, James C Phillips, Alexander D MacKerell Jr, Klaus Schulten, and Benoît Roux. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical drude oscillators in namd. *The journal of physical chemistry letters*, 2(2):87–92, 2010.

[43] Laxmikant Kalé, Robert Skeel, Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten. Namd2: greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 151(1):283–312, 1999.

[44] Erwin Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical review*, 28(6):1049, 1926.

[45] Max Born and Robert Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[46] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[47] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[48] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[49] John P Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, 33(12):8822, 1986.

[50] Yoseph Imry and Rolf Landauer. Conductance viewed as transmission. *Reviews of Modern Physics*, 71(2):S306, 1999.

[51] Scheer Elke and Cuevas Juan Carlos. *Molecular electronics: an introduction to theory and experiment*, volume 15. World Scientific, 2017.

[52] Joseph Jacobson, Gunnar Björk, Isaac Chuang, and Yoshihisa Yamamoto. Photonic de broglie waves. *Physical review letters*, 74(24):4835, 1995.

[53] William A Goddard III, Donald Brenner, Sergey Edward Lyshevski, and Gerald J Iafrate. *Handbook of nanoscience, engineering, and technology*. CRC press, 2007.

[54] Yigal Meir and Ned S Wingreen. Landauer formula for the current through an interacting electron region. *Physical review letters*, 68(16):2512, 1992.

[55] Supriyo Datta. *Electronic transport in mesoscopic systems*. Cambridge university press, 1997.

[56] Thomas Frederiksen, Magnus Paulsson, Mads Brandbyge, and Antti-Pekka Jauho. Inelastic transport theory from first principles: Methodology and application to nanoscale devices. *Physical Review B*, 75(20):205413, 2007.

[57] Stefano Baroni, Stefano De Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Reviews of Modern Physics*, 73(2):515, 2001.

[58] Jing-Tao Lü, Rasmus B Christensen, Giuseppe Foti, Thomas Frederiksen, Tue Gunst, and Mads Brandbyge. Efficient calculation of inelastic vibration signals in electron transport: Beyond the wide-band approximation. *Physical Review B*, 89(8):081405, 2014.

[59] Tadahiro Komeda. Chemical identification and manipulation of molecules by vibrational excitation via inelastic tunneling process with scanning tunneling microscopy. *Progress in Surface science*, 78(2):41–85, 2005.

[60] Magnus Paulsson, Thomas Frederiksen, Hiromu Ueba, Nicolás Lorente, and Mads Brandbyge. Unified description of inelastic propensity rules for electron transport through nanoscale junctions. *Physical review letters*, 100(22):226604, 2008.

[61] BC Stipe, MA Rezaei, and W Ho. Single-molecule vibrational spectroscopy and microscopy. *Science*, 280(5370):1732–1735, 1998.

[62] Robert C Jaklevic and John Lambe. Molecular vibration spectra by electron tunneling. *Physical Review Letters*, 17(22):1139, 1966.

[63] Magnus Paulsson, Thomas Frederiksen, and Mads Brandbyge. Modeling inelastic phonon scattering in atomic-and molecular-wire junctions. *Physical Review B*, 72(20):201101, 2005.

[64] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[65] Scott E Feller and Alexander D MacKerell. An improved empirical potential energy function for molecular simulations of phospholipids. *The Journal of Physical Chemistry B*, 104(31):7510–7515, 2000.

[66] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[67] Yun Luo and Benoît Roux. Simulation of osmotic pressure in concentrated aqueous salt solutions. *The Journal of Physical Chemistry Letters*, 1(1):183–189, 2009.

[68] Aleksij Aksimentiev and Klaus Schulten. Imaging α-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophysical journal*, 88(6):3745–3761, 2005.

[69] Mikhail A Lomize, Andrei L Lomize, Irina D Pogozheva, and Henry I Mosberg. Opm: orientations of proteins in membranes database. *Bioinformatics*, 22(5):623–625, 2006.

[70] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[71] Pierre Thevenet, Yimin Shen, Julien Maupetit, Frederic Guyon, Philippe Derreumaux, and Pierre Tuffery. Pep-fold: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic acids research*, 40(W1):W288–W293, 2012.

[72] James Gumbart, Fatemeh Khalili-Araghi, Marcos Sotomayor, and Benoît Roux. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1818(2):294–302, 2012.

[73] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The siesta method for ab initio order-n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

[74] Mads Brandbyge, José-Luis Mozos, Pablo Ordejón, Jeremy Taylor, and Kurt Stokbro. Density-functional method for nonequilibrium electron transport. *Physical Review B*, 65(16):165401, 2002.

[75] Richard M Martin. *Electronic structure: basic theory and practical methods*. Cambridge university press, 2004.

[76] Hendrik J Monkhorst and James D Pack. Special points for brillouin-zone integrations. *Physical review B*, 13(12):5188, 1976.

[77] Norman Troullier and José Luís Martins. Efficient pseudopotentials for plane-wave calculations. *Physical review B*, 43(3):1993, 1991.

[78] Shuanglong Liu, Argo Nurbawono, and Chun Zhang. Density functional theory for steady-state nonequilibrium molecular junctions. *Scientific reports*, 5:15386, 2015.

[79] ChiYung Yam, Xiao Zheng, GuanHua Chen, Yong Wang, Thomas Frauenheim, and Thomas A Niehaus. Time-dependent versus static quantum transport simulations beyond linear response. *Physical Review B*, 83(24):245448, 2011.

[80] Kamal K Saha, Marija Drndic, and Branislav K Nikolic. Dna base-specific modulation of microampere transverse edge currents through a metallic graphene nanoribbon with a nanopore. *Nano letters*, 12(1):50–55, 2011.

[81] Jariyanee Prasongkit, Anton Grigoriev, Biswarup Pathak, Rajeev Ahuja, and Ralph H Scheicher. Transverse conductance of dna nucleotides in a graphene nanogap from first principles. *Nano letters*, 11(5):1941–1945, 2011.

[82] Zuanyi Li, Haiyun Qian, Jian Wu, Bing-Lin Gu, and Wenhui Duan. Role of symmetry in the transport properties of graphene nanoribbons under bias. *Physical review letters*, 100(20):206802, 2008.

[83] Jariyanee Prasongkit, Gustavo T Feliciano, Alexandre R Rocha, Yuhui He, Tanakorn Osotchan, Rajeev Ahuja, and Ralph H Scheicher. Theoretical assessment of feasibility to sequence dna through interlayer electronic tunneling transport at aligned nanopores in bilayer graphene. *Scientific Reports*, 5:17560, 2015.

[84] Gustavo T Feliciano, Carlos Sanz-Navarro, Mauricio Domingues Coutinho-Neto, Pablo Ordejón, Ralph H Scheicher, and Alexandre Reily Rocha. Capacitive dna detection driven by electronic charge fluctuations in a graphene nanopore. *Physical Review Applied*, 3(3):034003, 2015.

[85] P Giannozzi. P. giannozzi, s. baroni, n. bonini, m. calandra, r. car, c. cavazzoni, d. ceresoli, gl chiarotti, m. cococcioni, i. dabo, a. dal corso, s. de gironcoli, s. fabris, g. fratesi, r. gebauer, u. gerstmann, c. gougoussis, a. kokalj, m. lazzeri, l. martin-samos, n. marzari, f. mauri, r. mazzarello, s. paolini, a. pasquarello, l. paulatto, c. sbraccia, s. scandolo, g. sclauzero, ap seitsonen, a. smogunov, p. umari, and rm wentzcovitch, j. phys.: Condens. matter 21, 395502 (2009). *J. Phys.: Condens. Matter*, 21:395502, 2009.

[86] Roger Fletcher. A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322, 1970.

[87] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry*, 27(15):1787–1799, 2006.

[88] Yanan Zhao, Brian Ashcroft, Peiming Zhang, Hao Liu, Suman Sen, Weisi Song, JongOne Im, Brett Gyarfas, Saikat Manna, Sovan Biswas, et al. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nature nanotechnology*, 9(6):466, 2014.

[89] Hannah L McFarland, Towfiq Ahmed, Jian-Xin Zhu, Alexander V Balatsky, and Jason T Haraldsen. First-principles investigation of nanopore sequencing using variable voltage bias on graphene-based nanoribbons. *The journal of physical chemistry letters*, 6(13):2616–2621, 2015.

[90] Mark A Reed, C Zhou, CJ Muller, TP Burgin, and JM Tour. Conductance of a molecular junction. *Science*, 278(5336):252–254, 1997.

[91] Syed Ghazi Sarwat, Pascal Gehring, Gerardo Rodriguez Hernandez, Jamie H Warner, G Andrew D Briggs, Jan A Mol, and Harish

Bhaskaran. Scaling limits of graphene nanoelectrodes. *Nano letters*, 17(6):3688–3693, 2017.

[92] Tchavdar N Todorov. Tight-binding simulation of current-carrying nanostructures. *Journal of Physics: Condensed Matter*, 14(11):3049, 2002.

[93] Gemma C Solomon, Carmen Herrmann, Thorsten Hansen, Vladimiro Mujica, and Mark A Ratner. Exploring local currents in molecular junctions. *Nature chemistry*, 2(3):223, 2010.

[94] T. Frederiksen, M. Paulsson, M. Brandbyge, and A.-P. Jauho. Inelastic transport theory from first principles: Methodology and application to nanoscale devices. *Phys. Rev. B*, 75(20):205413, 2007.

[95] Aldo Eugenio Rossini and Giuseppe Zollo. Chirality dependence of the tunneling current in graphene nano-gaps for peptide sequencing. In *AIP Conference Proceedings*, volume 1990, page 020003. AIP Publishing, 2018.

[96] AA Zamyatnin. Protein volume in solution. *Progress in biophysics and molecular biology*, 24:107–123, 1972.

[97] Wei Si and Aleksei Aksimentiev. Nanopore sensing of protein folding. *ACS nano*, 11(7):7091–7100, 2017.

[98] S Bonella, D Raimondo, E Milanetti, A Tramontano, and G Ciccotti. Mapping the hydropathy of amino acids based on their local solvation structure. *The Journal of Physical Chemistry B*, 118(24):6604–6613, 2014.

[99] Alina Asandei, Aldo E Rossini, Mauro Chinappi, Yoonkyung Park, and Tudor Luchian. Protein nanopore-based discrimination between selected neutral amino acids from polypeptides. *Langmuir*, 33(50):14451–14459, 2017.

[100] Mauro Chinappi and Fabio Cecconi. Protein sequencing via nanopore based devices: a nanofluidics perspective. *Journal of Physics: Condensed Matter*, 30(20):204002, 2018.

[101] Aldo Eugenio Rossini, Fabrizio Gala, Mauro Chinappi, and Giuseppe Zollo. Peptide bond detection via graphene nanogaps: a proof of principle study. *Nanoscale*, 10(13):5928–5937, 2018.

[102] Emilio Artacho, Eduardo Anglada, Oswaldo Diéguez, Julian D Gale, Alberto García, Javier Junquera, Richard M Martin, Pablo Ordejón, José Miguel Pruneda, Daniel Sánchez-Portal, et al. The siesta method; developments and applicability. *Journal of Physics: Condensed Matter*, 20(6):064208, 2008.

[103] Pablo Ordejón, Emilio Artacho, and José M Soler. Self-consistent order-n density-functional calculations for very large systems. *Physical Review B*, 53(16):R10441, 1996.

[104] Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters*, 73(2):393–398, 1980.

[105] Massimiliano Di Ventra and Sokrates T Pantelides. Hellmann-feynman theorem and the definition of forces in quantum time-dependent and transport problems. *Physical Review B*, 61(23):16207, 2000.